

**HFS MAJOR APPLIED RESEARCH PAPER
NO. 2**

**TECHNICAL AND ECONOMIC EFFICIENCY
IN THE PRODUCTION OF HEALTH SERVICES**

**Phase 1:
Review of Concepts and Literature,
and Preliminary Field Work Design**

Submitted to:

Health Services Division
Office of Health
Bureau for Research and Development
Agency for International Development

By

Ricardo BITRAN, Ph.D.
Abt Associates Inc.

December 1992

Health Financing and Sustainability (HFS) Project
Abt Associates Inc., Prime Contractor
4800 Montgomery Lane, Suite 600
Bethesda, MD 20814 USA
Tel: (301) 913-0500 Fax: (301) 652-7791
Telex: 312636

The Urban Institute, Subcontractor
Management Sciences for Health, Subcontractor

A.I.D. Contract No. DPE-5974-Z-00-9026-00

ABSTRACT

This report completes the first phase of the HFS Project's three-phase major applied research project entitled "Private Sector: Private and Public Differences in Efficiency." This first phase lays the groundwork for phase two (field work) and phase three (analysis) of the research project. This paper provides a selective review of studies of health service production efficiency that have been performed in developed and developing countries. The review examines techniques used to measure efficiency, methodological problems associated with these measurements, and possible solutions. It also provides information about the efficiency of public and private health services in developing countries. Finally, the report describes opportunities for phase two field work, and discusses efficiency measurement techniques for data analysis. The field work opportunities described are: 1) a comparative study of hospital efficiency in Ecuador for five different types of providers, and 2) a study of productive efficiency for private and public facilities in Senegal, including facilities from various levels.

TABLE OF CONTENTS

LIST OF EXHIBITS	v
ACRONYMS	vi
ACKNOWLEDGEMENTS	vii
EXECUTIVE SUMMARY	viii
1.0 INTRODUCTION	10
2.0 DEFINITIONS: EFFICIENCY, EQUITY, AND QUALITY IN HEALTH CARE	12
2.1 QUALITY OF CARE	12
2.2 TECHNICAL EFFICIENCY	13
2.3 ECONOMIC EFFICIENCY	15
3.0 MEASUREMENT OF TECHNICAL AND ECONOMIC EFFICIENCY IN THE PRODUCTION OF HEALTH CARE	16
3.1 EFFICIENCY AND QUALITY OF CARE	16
3.2 EFFICIENCY AND CASE MIX	17
3.3 ECONOMIC EFFICIENCY AND INPUT PRICE ISSUES	18
3.4 TECHNICAL AND ECONOMIC EFFICIENCY: THEIR RELATIVE IMPORTANCE	20
4.0 REVIEW OF SELECTED STUDIES ON HEALTH SERVICES EFFICIENCY FROM THE DEVELOPED WORLD	22
4.1 THE PROBLEM OF QUALITY HETEROGENEITY	22
4.2 THE PROBLEM OF CASE-MIX DIFFERENCES	25
4.3 THE MEASUREMENT OF EFFICIENCY AND INPUT PRICE ISSUES ..	30
4.4 MEASUREMENT OF EFFICIENCY: RATIO ANALYSIS	30
4.5 MEASUREMENT OF EFFICIENCY: ECONOMETRIC ANALYSIS OF HEALTH FACILITY COSTS	31
4.6 MEASUREMENT OF TECHNICAL EFFICIENCY: DATA ENVELOPMENT ANALYSIS AND RELATED TECHNIQUES	37
4.7 CONCLUSIONS	44
5.0 REVIEW OF SELECTED STUDIES ON HEALTH SERVICES EFFICIENCY FROM THE DEVELOPING WORLD	46
5.1 LITERATURE REVIEW	46
5.2 SUMMARY AND CONCLUSIONS	52
6.0 PRELIMINARY RESEARCH DESIGN FOR PHASE TWO FIELD WORK	54
6.1 A COMPARATIVE STUDY OF EFFICIENCY IN GOVERNMENT AND PRIVATE INPATIENT FACILITIES IN ECUADOR	54

6.1.1	Assessment of Overall Hospital Technical Efficiency Using the DEA Technique	54
6.1.1.1	Study Goals and Objectives	54
6.1.1.2	Method and Workplan	55
6.1.1.3	Level of Effort	58
6.1.2	Assessment of Economic Efficiency for a Selected Sample of Health Interventions	58
6.1.2.1	Study Goals and Objectives	58
6.1.2.2	Method and Workplan	60
6.1.2.3	Level of Effort	63
6.2	A COMPARATIVE STUDY OF EFFICIENCY IN GOVERNMENT AND PRIVATE INPATIENT FACILITIES IN SENEGAL	65
6.2.1.	Goals and Objectives	65
6.2.2.	Method and Workplan	65
6.2.3.	Level of Effort	69
BIBLIOGRAPHY		71

LIST OF EXHIBITS

Exhibit 1:	Technically Efficient Production Possibilities Frontier or Isoquant	14
Exhibit 2:	Technical and Economic Efficiency	15
Exhibit 3:	Technical Quality of Care and Efficiency	16
Exhibit 4:	Case Mix and Efficiency	18
Exhibit 5:	Economic Efficiency and Input Prices	19
Exhibit 6:	Social Economic Inefficiency in Production Due to an Undervalued Input	20
Exhibit 7:	Technical Versus Economic Efficiency	21
Exhibit 8:	Approaches for Standardizing Hospital Output According to Tatchell (1983)	22
Exhibit 9:	A Graphical Interpretation of Technical and Economic Inefficiency	38
Exhibit 10:	Several Providers Producing Different Quantities of Single, Homogeneous Output (Q) with Two Inputs (X_a and X_b)	39
Exhibit 11:	Efficiency Points and Unit Isoquant (adapted from Charnes et al, 1978)	40
Exhibit 12:	Study of Hospital Efficiency in Ecuador Using DEA Technique: Gantt Chart	56-57
Exhibit 13:	Study of Hospital Efficiency in Ecuador Using DEA Technique: Estimated Level of Effort	59
Exhibit 14:	Study of Hospital Efficiency in Ecuador for Selected Sample of Interventions: Gantt Chart	61-62
Exhibit 15:	Study of Hospital Efficiency in Ecuador for Selected Sample of Interventions: Estimated Level of Effort	64
Exhibit 16:	Study of Health Facility Costs and Efficiency and Health Committee Performance in Senegal: Gantt Chart	66-67
Exhibit 17:	Study of Health Facility Costs and Efficiency and Health Committee Performance in Senegal: Level of Effort	70

LIST OF ACRONYMS

CER	Cost-Effectiveness Ratio
DEA	Data Envelopment Analysis
DMU	Decision Making Unit
EGLS	Estimated Generalized Least Squares
FTE	Full-Time Equivalent
IHU	Intermediate Health Unit
ITE	Index of Technical Efficiency
LOS	Length of Stay
MOH	Ministry of Health
OLS	Ordinary Least Squares
WLS	Weighted Least Squares

ACKNOWLEDGEMENTS

This paper benefitted from comments by HFS staff at a seminar held in December 1991, from written comments by Marty Makinen, John Tilney, Jr., Avi Dor, Philip Musgrove, and Taryn Vian, from reviews by Carl Stevens and David Gwatkin, and from suggestions by the members of the HFS Technical Advisory Group during the 1992 meeting. I thank Dayl Donaldson for bringing to my attention several useful references on health service efficiency measurement.

EXECUTIVE SUMMARY

This paper constitutes the first phase of a three-phase HFS major applied research project in the area of "Private Sector: Private and Public Differences in Efficiency" (HFS Applied Research Agenda, 1991). The document provides the foundations for phases two (field work) and three (analysis). A companion HFS phase one paper entitled "Provider Incentives and Productive Efficiency in Government Health Services" (Bitran and Block, 1992) complements the definitions and analysis of this document.

The purpose of this paper is to review studies of health service production efficiency conducted in industrialized countries as well as in developing nations. Through this review, HFS seeks both to improve its understanding of available techniques for the measurement of efficiency, and to gain empirical knowledge about the efficiency of public and private health services in developing countries. Increased knowledge about public-private levels and differences in efficiency can contribute to the public policy debate on how to improve the efficiency with which governments spend their scarce resources for the provision of health care to their populations.

The paper describes techniques for the measurement of technical and economic efficiency in the production of health services, the methodological problems associated with these measurements, and possible solutions; it also provides a selective review of efficiency studies done in developing and developed countries.

The final section identifies field work opportunities for phase two. A preliminary discussion of efficiency measurement techniques appropriate for data analysis is presented. The field work opportunities are:

- A comparative study of hospital efficiency in Ecuador for five different types of providers; and
- A study of productive efficiency for private and public facilities in Senegal, including facilities from various levels.

The study in Ecuador seeks to assess the efficiency of hospitals run by the Ministry of Public Health, the Junta de Beneficencia, the Social Security Administration, the private sector, and the military. One hospital of each type would be included in the study. USAID/Quito has recently asked HFS to develop a research proposal. Possible measurement methods for this study are Data Envelopment Analysis; assessment of efficiency for a small selected sample of interventions and for a probabilistic sample of patients, controlling for quality of care; and ratio analysis.

In Senegal, USAID/Dakar has requested that HFS perform a study of health facility efficiency using a sample of about 120 Ministry of Health and private facilities, including hospitals, maternities, health centers, posts, and huts. Efficiency measurement studies that can be used in this study for the case of hospitals include ratio analysis and assessment of efficiency using selected interventions and a sample of patients. For the ambulatory facilities, DEA and econometric

estimation of cost and production functions, and the derivation of associated measures of efficiency are proposed.

1.0 INTRODUCTION

Improving efficiency in government-run health services of developing countries is a major goal of governments and donors alike. Higher efficiency can allow greater production and better quality of services without consuming further financial or real resources.

Numerous strategies have been advanced by experts to improve the efficiency of government-run health services. One strategy is the outright privatization of government-owned operations. It is based on the notion that private ownership entails mechanisms which promote economic efficiency, while public ownership lacks such mechanisms. Despite its popularity, this policy recommendation is based on weak empirical foundations, insofar as the health sector is concerned. While inefficiencies in government-run health services have been documented in the developing country context¹, to date only a handful of studies have compared efficiency between public and private providers.² For the most part, these studies have based their results on rather small sample sizes and have been constrained by poor data.³ Of course, if greater private efficiency were found, appropriate financing and management systems would have to be put into place to ensure that the superior efficiency of private provision were in fact captured by the public. The study of such systems should be the subject of a separate paper.

The purpose of this paper is to review studies of health service production efficiency conducted in industrialized countries as well as in developing nations. Through this review, HFS seeks both to improve its understanding of available techniques for the measurement of efficiency, and to gain empirical knowledge about the efficiency of public and private health services in developing countries. Increased knowledge about public-private levels and differences in efficiency can contribute to the public policy debate on how to improve the efficiency with which governments spend their scarce resources for the provision of health care to their populations.

The paper presents a review of selected studies on productive efficiency from industrialized and developing countries. For the industrialized country literature, the emphasis of the review is on methods rather than on findings; for the developing country literature, emphasis is put on both the methods and the findings. This review is by no means comprehensive; rather, it is selective and has the purpose of highlighting methodological problems and approaches and describing the various measurement techniques available.⁴

¹ See, for example, Wouters (1990), Bitran and Dunlop (1989), and Lewis et al. (1990).

² Examples are Rodríguez and Jiménez (1985) and Wouters (1990). For a review of developing country studies on health facility costs and efficiency, see Barnum and Kutzin (1990).

³ For example, the data sets used by Bitran and Dunlop (1989) and by Wouters (1990) did not include information about prices of production inputs.

⁴ For a review of advances in econometrics and other mathematical techniques regarding the general problem of measuring productive efficiency in firms, see the 1980 and 1990 Supplements to the Journal of Econometrics.

This paper focuses on the measurement of efficiency of individual health care facilities or providers. It does not address issues of overall health system efficiency, which involve assessing the efficiency of an array of providers at various levels of the health care delivery system. For a discussion of health system efficiency, see the work of Nutting et al. (1981 and 1982).

This document constitutes the first phase of a three-phase major applied research initiative of the HFS Project. It falls under the research topic of "Public Sector: Public and Private Differences in Efficiency."⁵ Phase 2 constitutes the field work (mainly data collection) and phase 3 is for data analysis.

The paper's intended audience includes health financing and economics researchers, graduate health economics students, and policymakers with a background in health economics.

The remainder of the document is organized as follows. Section 2 provides definitions of basic efficiency-related concepts. Section 3 is a discussion of efficiency measurement issues. A review of selected studies on health facility efficiency that use data from developed countries is provided in Section 4. A similar review for developing countries is presented in Section 5. Finally, a summary of the findings from the reviews and a preliminary discussion of the field work design are presented in Section 6.

⁵ See HFS Applied Research Agenda, HFS Project, 1991.

2.0 DEFINITIONS: EFFICIENCY, EQUITY, AND QUALITY IN HEALTH CARE

Before proceeding to an analysis of efficiency measurement, it is necessary to clarify the underlying concepts around which this discussion revolves: quality of care, technical efficiency, and economic efficiency.

2.1 QUALITY OF CARE

Quality of care is of critical concern in the analysis of health care systems. This concern is due to the fact that health services can vary widely in quality, whereas goods and services in other markets, particularly in commodity markets, tend to exhibit greater quality homogeneity. Also, quality of care is not easily discerned by consumers of health services, because such services are not always traded in competitive markets. Additionally, quality of care affects health care demand.

When studying health services, it is useful to distinguish between *technical quality* and *perceived quality* of care. The definition of technical quality adopted here is that used by Wyszewianski et al.(1987), who consider technical quality for care A to be higher than that for care B if care A is likely to make a greater net contribution to the patient's health and well-being than care B. These authors further explain that the expected effect on health is determined in part by whether one type of care is more appropriate than another, and in part by whether the procedure is correctly performed.

Quality also has a more subtle dimension relating to the perceptions of consumers of health care. Demand for health care is, in part, a function of perceived quality. While, ultimately, this perception may be formed by whether or not the patient's condition improves, perceptions of quality may also be colored by whether or not drugs are prescribed, or by how politely the patient is treated by facility staff.

Barnum and Kutzin (1990) recognize that the quality of health care has two dimensions: the medical or technical dimension and the consumer's dimension.

Quality has both supply and demand characteristics. The critical demand issue is "perceived" quality: the consumer's assessment of the relative quality of different health care providers... Adequate staff and supplies are obvious supply side factors affecting "actual" quality of services that are important in affecting perceived quality. [Chapter 3, p.17]

Technical quality and perceived quality are not competing but rather complementary definitions of health care quality. Technical quality is a concept that encompasses the medical procedures performed and their effects on patients' health. Perceived quality influences the consumers' decisions to seek medical care and to choose a particular provider. Health analysts and policymakers must be concerned with both dimensions of quality, since both are important in understanding health care systems.

Two additional central concepts, pertaining to health care production and quality, need to be defined at this point: *health care output* and *health care outcome*. Health output denotes the number of units of medical services or procedures performed. Examples of output measures are: number of hospital discharges, number of surgical interventions, number of children immunized, and number of curative outpatient consultations. Health outcomes, in contrast, are measures of the population's health status. The improvement of health outcomes is the ultimate goal of health interventions; the provision of health outputs is intended to have a positive effect on people's health outcomes. The number of healthy days of life saved, deaths averted, and illnesses averted are all common examples of health outcome measures, as they relate to health interventions.

There is a direct link between the concept of technical quality and the notion of health outcome. In fact, technical quality can only be gauged fully when the effect of health services on health outcomes can be established. Unfortunately, health service professionals and researchers often have at their disposal information about health outputs, from facility statistics on services provided, while information about outcomes is more rare. Further, establishing a relationship between health output levels and health outcome levels is a complex empirical task, because health status is influenced by a large number of variables, such as nutrition, hygiene, health care, and life style; isolating the effect of each variable, including the consumption of health outputs, on health is thus not a straightforward exercise.

As is explained in the next section, knowledge about health care quality, and thus outcomes, is essential to the analysis of health services efficiency. In practice, however, the empirical obstacles faced by researchers in linking specific health outputs with outcomes, means that they often have to limit themselves to expressing efficiency as a function of the more easily observable health outputs.

2.2 TECHNICAL EFFICIENCY

Two important concepts intervene in the analysis of efficiency of a production process: *technical efficiency* and *economic efficiency*.⁶ A procedure is technically efficient if production inputs (e.g., labor, drugs, equipment) are combined in a way that yields the maximum feasible output (e.g., outpatient visits, hospitalizations).⁷

⁶ Health service researchers in the U.S. use the terms "efficacy" to refer to technical efficiency and "appropriateness" to denote economic efficiency.

⁷ Pauly (1970), p.114

In microeconomic terms, a technically efficient production process is one that is placed along the *production possibilities frontier* or *isoquant*. This is exemplified in Exhibit 1, for a simple medical production process that uses only two production inputs, X_a and X_b (these inputs can be viewed as being doctor and nurse time, or doctor time and drugs, etc.).⁸ Any point along the production possibilities frontier QQ represents a technically efficient way of combining various quantities of production inputs X_a and X_b to produce the same amount of output Q . For example, while points 1 and 2 differ in the combination of X_a and X_b (production at 1 is more intensive in X_b than at 2), both permit production of the same quantity Q . Points 1 and 2, like all other points on the frontier QQ

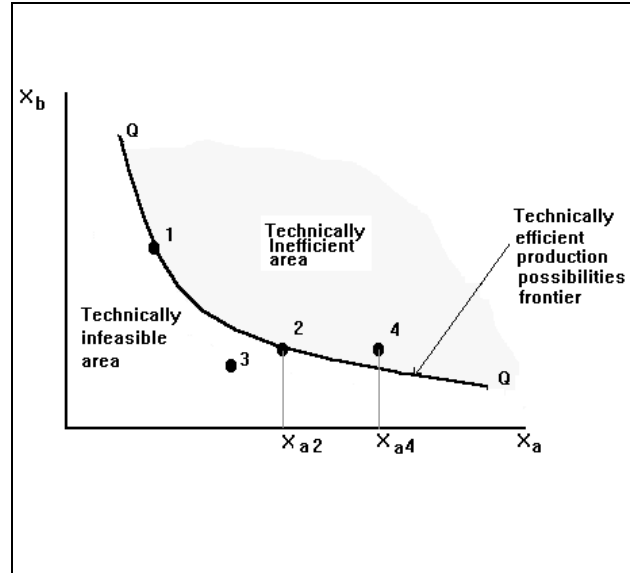


Exhibit 1: Technically Efficient Production Possibilities Frontier or Isoquant

are technically efficient because it is not possible to produce Q with smaller quantities of either X_a or X_b , as depicted by the line (there is no room for further gain in technical efficiency). Point 3, like all points to the left of the production possibilities frontier, is infeasible: any reduction in the amounts of X_a and X_b from the amounts represented by the frontier necessarily translates into a drop in Q . In contrast, point 4, like all points to the right of the production possibilities frontier, constitutes a technically inefficient way of producing Q : technical efficiency can be improved by moving production from 4 to 2, thereby reducing the amount of X_a from X_{a4} to X_{a2} . In sum, one procedure is considered more technically efficient than another, either if it produces the same quantity of output using fewer inputs, or if it produces a greater quantity of outputs using the same resources.

The production possibilities frontier represents all the possible combinations of inputs which permit production of the same quantity of health care output. It will be further assumed that technical quality of care also remains constant along the production possibilities frontier. Thus, not only does any combination of inputs X_a and X_b along the curve permit production of quantity Q of medical care output, but also any such combination delivers medical care of constant technical quality, i.e., with the same effect on patients' health status.

Perceived quality of care is not necessarily constant along the frontier QQ , however. For example, while both points 1 and 2 in Exhibit 1 permit production of Q with the same impact on patients' health, some patients may find one point superior, or of better perceived quality, than the

⁸While most health care production processes employ more than two inputs, the visual illustrations used in this document are two-dimensional, for graphical convenience. The definitions and analyses presented here are directly generalizable to multiple input production processes.

other. For instance, if X_a is doctor time and X_b is nurse time, some patients may prefer point 2 because it uses more doctor time than point 1, which is more intensive in nurse time.

2.3 ECONOMIC EFFICIENCY

Economic efficiency extends the concept of technical efficiency to take into account the relative prices of production inputs. A procedure is economically efficient if inputs are combined to produce a given level of output *at minimum cost*. In general, while many technically efficient alternatives might present themselves to produce a given quantity Q , there is only *one* economically efficient way of doing so.⁹

Exhibit 2 helps to illustrate the fundamental difference and relationship between technical and economic efficiency. Suppose that the unit prices of production inputs X_a and X_b are W_a and W_b , respectively. If the health facility is allocated a budget B_1 , then line B_1 represents the facility's budget constraint. The constraint is given by the equation: $B_1 = X_a \cdot W_a + X_b \cdot W_b$. Any point along the budget constraint line, such as points 1 and 3, consumes the total budget B_1 . However, point 1 is preferable to 3 because at 1 quantity Q is produced, whereas at point 3 the smaller quantity Q'' is produced. Further, of all the technically efficient points along the frontier QQ , point 1 is the most economically efficient way of producing quantity Q . Point 2 is technically as efficient as 1, but is less economically efficient, since production at 2 requires a budget of B_2 , higher than B_1 . Graphically, the economically efficient point (point 1) corresponds to the tangency between the budget constraint and the production possibilities frontier.

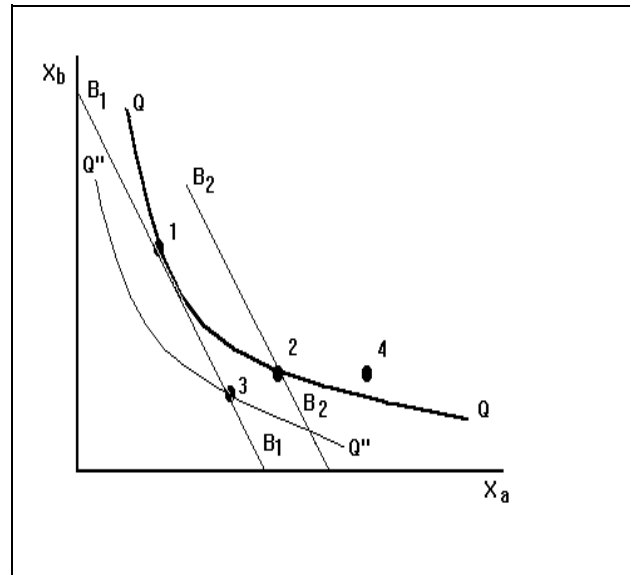


Exhibit 2: Technical and Economic Efficiency.

With the above definitions in mind, we turn to a discussion of efficiency measurement issues.

⁹ There are some unusual production processes which will display more than one economically efficient configuration.

3.0 MEASUREMENT OF TECHNICAL AND ECONOMIC EFFICIENCY IN THE PRODUCTION OF HEALTH CARE

The purpose of this section is to provide a conceptual discussion about issues in the measurement of technical and economic efficiency in the production of health services. The discussion begins with an analysis of efficiency measurement and quality of care issues. Next, I explain how efficiency measurement for a group of providers becomes a more complex problem when their case mix varies. An analysis of efficiency measurement and input price issues follows. The section concludes with a brief discussion about the use and importance of technical and economic efficiency.

3.1 EFFICIENCY AND QUALITY OF CARE

Comparisons of technical and economic efficiency among providers must take into account their technical quality of care. Different levels of quality often consume different levels of production inputs. Unless quality differences are considered, the researcher seeking to measure efficiency of a group of providers may obtain a distorted picture.

To illustrate this point, consider the two providers of Exhibit 3, D and E, each capable of producing the same volume of output (e.g., Q ambulatory visits) according to their respective production possibilities frontiers. While both providers operate at the same output level, they produce care of different technical quality: provider D is assumed to provide care of greater technical quality, H_1 , while provider E is supposed to produce care of a lower technical quality, H_2 .

Suppose that provider D operates at point 1 and provider E operates at point 2. If the researcher attempting to compare technical and economic efficiency between the two providers did not take into account their differences in technical quality, he would reach the conclusion that provider E is technically and economically more efficient than D. His conclusion would arise from the fact that provider E uses fewer production inputs than D (X_{a2} and X_{b2} versus X_{a1} and X_{b1} , respectively) and, as a consequence, provider E produces the level of output Q at a lower total cost than D. His conclusion would be wrong.

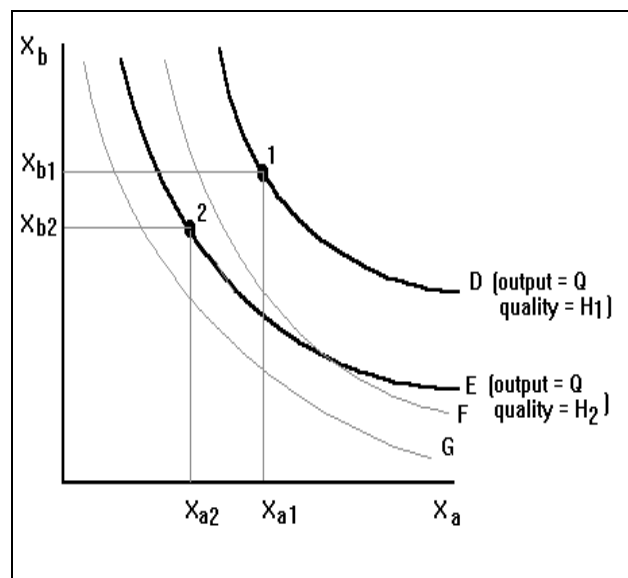


Exhibit 3: Technical Quality of Care and Efficiency

An appropriate comparison of efficiency is one which, at any given level of output, relates technical quality to input use. The researcher should therefore establish a relationship between H_1 and (X_{a1}, X_{b1}) for provider D and compare it with the equivalent relationship between H_2 and

(X_{a2}, X_{b2}) for the provider E. However, because the input levels X_a and X_b are in physical units, it is difficult to establish a quantitative relationship between health outcome levels (H) and the input levels.

A solution to that problem is the cost effectiveness technique, which consists of multiplying each production input by its price to compute the total cost of production, and then relating the total cost to the health outcome level.¹⁰ There exist other techniques for measuring and comparing technical and economic efficiency among various providers. Such techniques, presented in subsequent sections of this paper, include econometric estimation of production and cost functions and the Data Envelopment Analysis method.

Higher technical quality does not necessarily imply greater use of inputs, as is suggested by isoquants D and E in Exhibit 3. There, it is assumed that technical quality is higher along the isoquant D than along E, and also that resource use is greater for D. That does not necessarily have to be the case in all situations. For example, consider production of quantity Q according to F. Provider F's technical quality could be higher than E's, with F using smaller quantities of inputs when both providers operate at the far right of their possibilities (that is, production that is intensive in resource X_a). Or technical quality of provider G could be greater than that of E at all points, yet with G consuming fewer inputs than E and thus being technically and economically more efficient.

3.2 EFFICIENCY AND CASE MIX

Just like differences in technical quality of care obscure comparisons of efficiency among providers, so do differences in case mix. Case mix is an important, yet hard to define, concept (see Tatchell (1983), Hornbrook (1982a), and Hornbrook (1982b)) through which researchers attempt to define hospital output. Definitions available involve some or all of the following information: facilities (or services) available, intermediate and final services actually provided, complexity of the cases treated, and patient characteristics (for example, age and gender). Everything else being constant, one would expect efficient providers with different case mix to use different levels of production inputs. For example, a facility with a greater proportion of complex cases should be expected to use more resources in production than an otherwise identical facility treating a set of patients with fewer severe cases.

¹⁰For example, if W_a and W_b were the purchase prices of the two inputs, then the cost effectiveness ratio (CER) of providers D and E would be computed as follows:

$$CER_D = \frac{W_a \cdot X_{a1} + W_b \cdot X_{b1}}{H_1}$$

$$CER_E = \frac{W_a \cdot X_{a2} + W_b \cdot X_{b2}}{H_2}$$

The CER ratio computes the cost per unit of health outcome achieved. When a choice is to be made among alternative providers or interventions, the one with the highest cost effectiveness ratio is chosen.

Unless case mix is considered, comparative studies of technical and economic efficiency among several providers will likely be wrong. To illustrate this point, consider in Exhibit 4 the case of two providers, L and M, with L treating high severity patients, such as children with severe dehydration from dysentery, and M treating low severity patients, like children with mild dehydration from dysentery. Highly dehydrated children may need to remain hospitalized for several days, often receiving intravenous feeding and rehydration, and requiring close attention by the facility staff. Children with mild dehydration can be sent home with instructions to the parents to feed them oral rehydration salts and with the appropriate treatment for dysentery. Suppose that provider L operated at point 3 to treat high severity cases while provider M operated at point 4 to treat the milder cases. If case severity were not taken into account in this case, the uninformed researcher would wrongly conclude that provider M, the one with the lowest input use, is the more technically and economically efficient. If case mix were considered, however, the researcher would observe that the provider consuming the greatest amount of resources happens to treat the most severe cases. Without further analysis he could not make any definitive statements about relative efficiency.

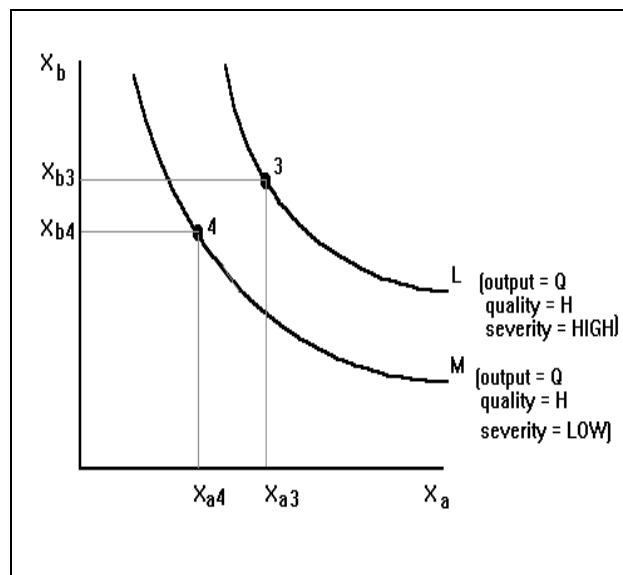


Exhibit 4: Case Mix and Efficiency

3.3 ECONOMIC EFFICIENCY AND INPUT PRICE ISSUES

Depending on a variety of circumstances, dealing with the incentives, constraints, and information available to health facility managers¹¹, some providers may operate in a technically efficient, yet economically inefficient manner. One such circumstance has to do with the prices of production inputs. Economic inefficiency arises when production occurs at a point that is not cost minimizing. This can happen either because facility managers do not know their input prices, or they know the prices but, for a number of reasons discussed elsewhere¹², they fail to minimize their costs.

¹¹ See the companion Phase 1 HFS paper on "Provider Incentives and Productive Efficiency", by Bitran and Block (1992).

¹² Ibid.

To distinguish between those two cases, consider the example of two providers operating at points 1 and 2 in Exhibit 5, each producing output level Q according to the same production possibilities frontier. Assume also that the two providers pay the same set of prices for their production inputs, X_a and X_b . Under those circumstances, provider 1 would be the most economically efficient of the two because its production cost would be B_1 , lower than B_2 . If the researcher wanting to study economic efficiency for these providers knew that both face the same input prices, he would not need to measure those prices at all to rightly conclude that 1 is more economically efficient than 2.

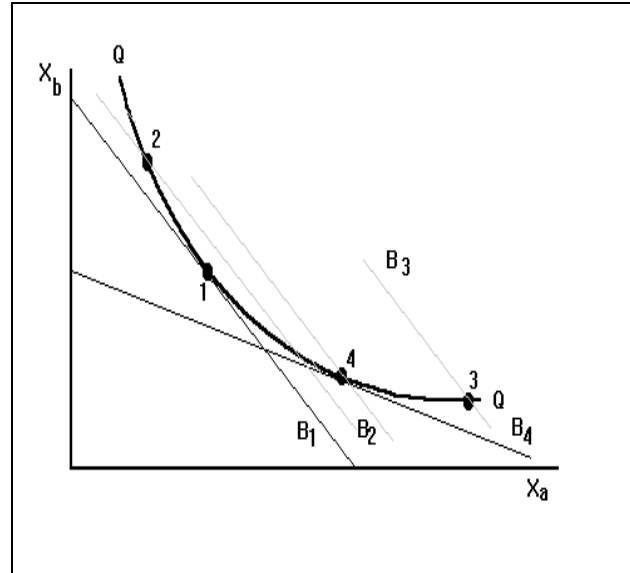


Exhibit 5: Economic Efficiency and Input Prices

Suppose, instead, that providers 1 and 2 face different input prices. Unless the researcher knew exactly what those sets of prices were, he would be unable to make any statements about the providers' relative economic efficiency. Both providers could be cost minimizers but, given the different prices they face, they would operate at different points along the production frontier. Or both could operate at points that are not cost minimizing. To ascertain their relative economic efficiency, knowledge and use of the price information would become essential.

In general, two types of circumstances, discussed above, can lead to economic inefficiency: technical inefficiency and technically efficient production that uses a mix of inputs that is not cost minimizing. There is a third cause of economic inefficiency, referred to as *social economic inefficiency* that can arise when the input prices faced by facility managers (for example, personnel wages or pharmaceutical products) depart from *social* (or *shadow*) prices. Social economic inefficiency is discussed next.

A socially undervalued production input will likely result in too much use of it relative to what a social economic evaluation would dictate, while an overvalued resource will result in too little use of it. For example, in a country with an overvalued currency, the transaction, or market price of imported inputs, such as pharmaceutical products, will be low relative to the shadow price of those goods. This may result in excessive use of drugs and medical supplies and, thus, in economic inefficiency from the viewpoint of society.

Exhibit 6 helps to illustrate, through an example, how divergences between transaction and social input prices can result in social economic inefficiency *even if* the provider is minimizing its private (as opposed to the social) costs of production. Suppose that the production of Q curative visits takes place at point 1, the economically efficient way of producing Q when the transaction input prices are W_a and W_b . The associated minimum provider cost of producing Q is equal to B_1 . Suppose, however, that one of the inputs is undervalued relative to the socially correct price. Assume, for example, that X_b is nurse time and W_b is the undervalued hourly wage of nurses. If the facility manager had to pay the higher, social wage to nurses, the budget constraint line would rotate downwards as depicted by the arrow in the figure. In that situation, the budget of B_1 would become insufficient to produce quantity Q . Production of Q at the minimum social cost would occur at point 2 at the higher total cost of B_2 . Notice that when the socially appropriate input prices are considered, then production at point 1 actually costs society B_3 , an amount greater than both B_1 and B_2 . Social economic efficiency would increase by moving production from point 1 to point 2 because the social cost of production would drop from B_3 to B_2 .

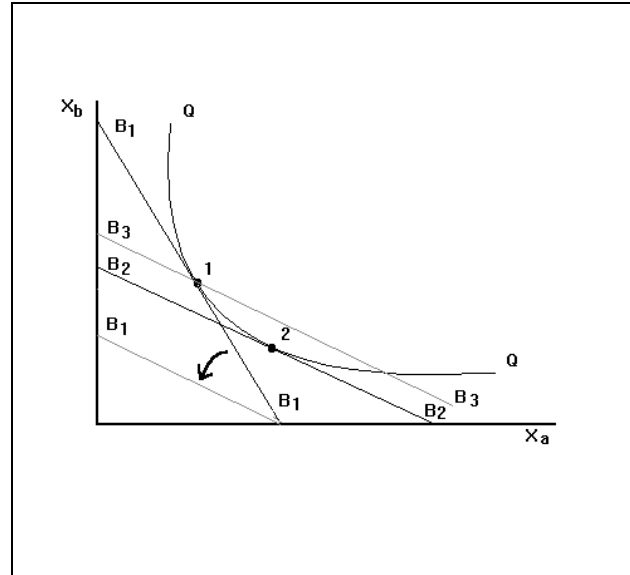


Exhibit 6: Social Economic Inefficiency in Production Due to an Undervalued Input

3.4 TECHNICAL AND ECONOMIC EFFICIENCY: THEIR RELATIVE IMPORTANCE

A question that has not been addressed so far is the relative importance of technical and economic efficiency: What type of efficiency should managers wishing to improve productive efficiency focus on: technical or economic efficiency?

As was suggested above, economic inefficiency often arises as a result of technical inefficiency. Thus, in many instances, improvements in technical efficiency (reducing the amount of resources used to produce a given quantity of output) will also result in greater economic efficiency (lowering the cost of producing a given quantity of output).

An example of this is shown in Exhibit 7. Technical and economic efficiency are improved when production moves from point 1 (at cost B_1) to point 2 (at cost $B_2 < B_1$). Unfortunately, greater technical efficiency will not always result in higher economic efficiency. In some instances, a gain in technical efficiency will result in a loss in economic efficiency. For example, production of Q according to point 1 is technically less efficient than according to point 3. Nevertheless, production at point 1 is economically more efficient than at point 3 (because $B_1 < B_3$). This is so because of the relative prices of inputs X_a and X_b .

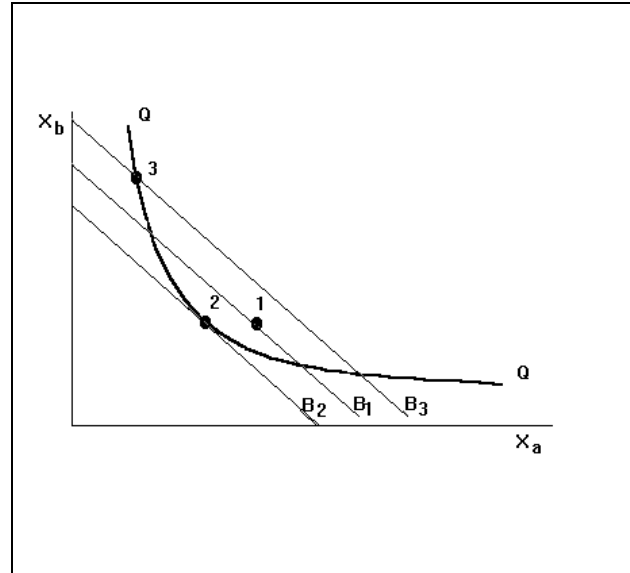


Exhibit 7: Technical Versus Economic Efficiency

Which of the two production points to choose, point 1 or point 3, will depend on the incentives and constraints of facility managers as well as on quality of care considerations. If managers were free to allocate production resources, and if they had the right cost minimizing incentives, then they should choose the lowest-cost, economically efficient production, or point 2. One factor which could possibly lead cost minimizing managers to choose a point along QQ other than point 2 would be the consumers' perceived quality of care. For example, suppose that consumers predominantly preferred point 3 to point 2, because they liked better those services that are intensive in resource X_b . Under those circumstances, if managers attempted to produce at point 2 they could find themselves with a quantity demanded smaller than Q , with unutilized resources, and thus with economic inefficiency. This discussion illustrates that the problem of deciding how to combine production inputs can become more complex when demand is taken into account, and it stresses the importance of understanding and considering demand information in the planning process. The ultimate solution to the problem will depend in part on how responsive demand is to alternative input configurations, as well as on the financial constraints of the facility, the goals of managers, and the mandate of the health system.

4.0 REVIEW OF SELECTED STUDIES ON HEALTH SERVICES EFFICIENCY FROM THE DEVELOPED WORLD

In recent years there has been a growing body of literature about the efficiency of health care services in industrialized countries, particularly in the U.S., where each year, over the past two decades, health care costs have risen much faster than the consumer price index. Knowledge about the levels and determinants of health services efficiency can help policymakers and health care managers determine whether there is inefficiency in production and, if so, to take measures aimed at curtailing costs while maintaining acceptable levels of quality and access.

This section presents a review of selected studies of health services efficiency conducted in industrialized countries.¹³ The purpose of the section is to discuss measurement issues and describe measurement techniques that may be applied to measuring efficiency in developing countries.

The section first illustrates how the problems discussed in Section 3 —quality and case-mix heterogeneity and price variations and distortions— are dealt with in the developed country literature. Next, the section presents alternative techniques for the measurement of efficiency, including: ratio analysis, econometric estimation of cost or production functions and associated measures of efficiency, and data envelopment analysis and related linear programming techniques.

4.1 THE PROBLEM OF QUALITY HETEROGENEITY

As already noted, quality variations in health services hinder efficiency comparisons among providers. Failure to control for quality differences may ascribe higher efficiency to lower-quality producers and vice-versa. Most studies of health facility costs and efficiency ignore the problem of quality variations, some implying that the only possible source of bias in efficiency measurement among providers is the difference in case-mix (see discussion below).¹⁴

Controlling for quality variations in the measurement of efficiency poses important methodological problems, mainly because of difficulties in measuring quality. Even if only a technical, or supply-side definition of quality is adopted, data limitations generally preclude an appropriate quality adjustment in studies of efficiency. If technical quality is to be measured on the basis of compliance with technical treatment standards and health outcomes, as in Wyszewianski

¹³ The studies on production efficiency and related studies reviewed for this section are: Vitaliano (1987), Caves and Christensen (1980), Sear (1991), Wyszewianski, Thomas, and Friedman (1987), Valdmanis (1990), Sherman (1984), Register and Bruning (1986), Eakin (1991), Eakin and Kniesner (1988), Reinhardt (1972), Goldman and Grossman (1983), Toda (1976), Byrnes and Valdmanis (1989), Golladay, Manser, and Smith (1974), Timmer (1970), Nutting, Burkhalter, Dietrick, and Helmick (1982), Nutting, Shorr, and Burkhalter (1981), Berwick (1989), Prescott, Cuthbertson, Fenwick, Garraway, and Ruckley (1978), Gold (1979), O'Donovan (1979), Lee, Wolfe, and Swenson (1984), Gururaj, Laude, Rajkumar, and Russo (1982), Pineault, Contandriopoulos, Valois, Bastian, and Lance (1985), Soulier and DeCoster (1982), Blandes, Culyer, and Walker (1987), Thorngren and Tinstman (1990), Bliersbach (1989), and Wan, Vanostenberg, Salley, Singley, and West (1987). For the reasons cited earlier, only a selected subset of these papers is discussed in this document.

¹⁴ See, for example, Register and Bruning (1987) and Lee et al. (1984).

et al. (1987), then a study of efficiency that controls for quality of care would have to obtain information about both the procedures performed and the patients' health status after the care is provided.

Collecting information on the extent to which medical standards are met in the provision of care can be time consuming, but is feasible within the realm of a study. Gathering information about health outcomes, however, can be a difficult task, particularly for medical procedures that are followed by slow patient recovery that takes place primarily outside of the health facility. Nevertheless, several studies have assessed health care quality by measuring clinical outcomes. Examples of studies which use the two alternative measures of technical quality are described below.¹⁵

In their study of one-day surgery in Quebec, Pineault et. al (1985) compare patient satisfaction, clinical outcomes, and costs of care between individuals undergoing one-day (ambulatory) surgery for tubal ligation, hernia repair, and meniscectomy, and individuals undergoing inpatient (overnight stay) surgery for those procedures. Patients participating in the study were randomly assigned to either one-day or inpatient surgery. This study combines both demand- and supply-side measures of quality (patient satisfaction and clinical outcome, respectively) with cost data.

Pineault et al. distinguish between "direct" and "indirect" measures of patient satisfaction, but adopt the latter, arguing that direct measures "are not specific enough and thus fail to discriminate between different modes of care."¹⁶ The indirect measures they retain pertain to "patients' perception [of the process of care]" and include:

(1) accessibility, as measured by perception of distance between home and hospital, controlling for real distance; (2) physician availability, as measured by at least one postoperative visit (excluding follow-up visits); and (3) patients' opinion concerning the appropriateness of the length of stay, and overall preference for the alternative mode of care [patients were asked if, given their actual experience, they would choose the same setting again or the alternative mode of care]. [p.173]

Evaluation of clinical outcomes came from two sources: the patients' view, which included seriousness of discomfort felt in the first 24 postoperative hours and the self-reporting of post-surgical problems (over a three-month follow-up period)¹⁷, and medical charts, which provided medical information on complications, general health status, symptoms, and complaints.

¹⁵ An example of this type of study done in the developing country context is described in Section 5 (see Shepard et al., 1991).

¹⁶ Pineault et al. (1985), p. 173. The authors point out that "direct measures of satisfaction are obtained by asking the patient to what extent he (she) is satisfied with various elements of the medical care process." (They also mention that there is a great deal of literature on this subject, and cite Ware et al. (1978) and Zastowny et al. (1983)).

¹⁷ In the case of meniscectomy patients, the recovery period extends beyond three months and is much longer than for the other surgical procedures. The study thus did not assess the long-term recovery for meniscectomy patients.

A weakness of the study is its inability to provide specific mechanisms for combining the three dimensions of the evaluation (patient satisfaction, clinical outcomes, and cost) into a unique performance measure which would allow comparisons between one-day and inpatient surgery. This is a problem if a procedure that is better on one dimension is worse on another.

A second limitation of this type of exercise is that the interpretation of the results requires value judgements about the relative importance of health outcomes. For example, more than 50% of one-day surgery patients would accept a repeat of their experience for tubal ligation and hernia repair. In addition, that form of care represents cost savings for both the patient and the Quebec health financing authority. From the above, the authors conclude that "tubal ligation and hernia repair seem to be surgical procedures for which one-day surgery is an appropriate form of care" [p.180]. This conclusion implicitly assigns a weight of zero to the other half of the patients who were dissatisfied and would not repeat their experience.

The Pineault et al. study is useful in illustrating the types of information that can be collected to assess qualitative aspects of health care but, like other studies of this nature, it provides little clue as to how qualitative measures can be combined with cost information for evaluative purposes.

The Pineault et al. study illustrates that quality of care can be assessed and controlled for in various ways when measuring efficiency. This includes measuring quality according to:

- (1) The degree of compliance with medical norms of practice;
- (2) Clinical outcomes; and
- (3) Various measures of patient satisfaction.

Combining two or three of these measures into a single quality indicator is impractical, for it requires value judgements that will generally be questionable. Separate use of any of these individual measures, however, in combination with information about costs, may shed light on truer efficiency differences among providers (i.e., differences less contaminated by quality variations). Nevertheless, two important problems remain unresolved. First, when two or more of these indicators are used, problems will arise if the quality measures yield conflicting results. For example, if provider A exhibits better compliance with clinical norms than provider B but patient satisfaction (or clinical outcome) is better with B than with A, quality differences between the two providers will remain ambiguous. A related, unresolved problem is the weighting of various quality indicators. An extreme example, that uses only clinical outcome as the quality measure, is a situation where the majority of provider A's patients recover fully but a small fraction die, versus the patients of provider B, all of whom recover only partially. In such cases, there will be ambiguities about quality differences between providers.

4.2 THE PROBLEM OF CASE-MIX DIFFERENCES

Comparative studies of health facility costs and efficiency have to overcome the difficult methodological problems arising from the wide diversity of health care activities produced and the

effects that this heterogeneity has on resource use, cost, and efficiency. This problem is particularly serious in the case of large secondary and tertiary hospitals.

Tatchell (1983) provides a comprehensive review of methods and studies used in the literature to tackle this problem in the context of hospital care. He identifies two main approaches for defining and measuring hospital output: one that uses "final outcomes" (or health outcomes or health status measures) and one that uses health output (services produced by the provider, such as number of discharges, patient days, number of deliveries, and so on). Tatchell argues, as has been done earlier in this paper, that the use of health outcome as a measure of health provider activity is hindered by two problems. First, health outcomes are difficult to measure since they often require patient follow-up over the long periods of time during which recovery is expected to occur. Second, even if changes in health outcomes can be measured, it is difficult to discern which part of the changes is attributable to health care and which is due to other factors affecting health status. The author therefore focuses his review on the use of health outputs (or "intermediate outcomes") to measure hospital output.

Tatchell argues that:

The main problem faced by researchers employing [health output] measures is lack of homogeneity. The problem presents itself in a number of ways and for a variety of reasons. A patient day of care, for example, is likely to differ:

- (a) between hospitals;
- (b) between different departments of the same hospital;
- (c) over time (both short term for a particular patient and in the longer term overall).

These differences may arise because of:

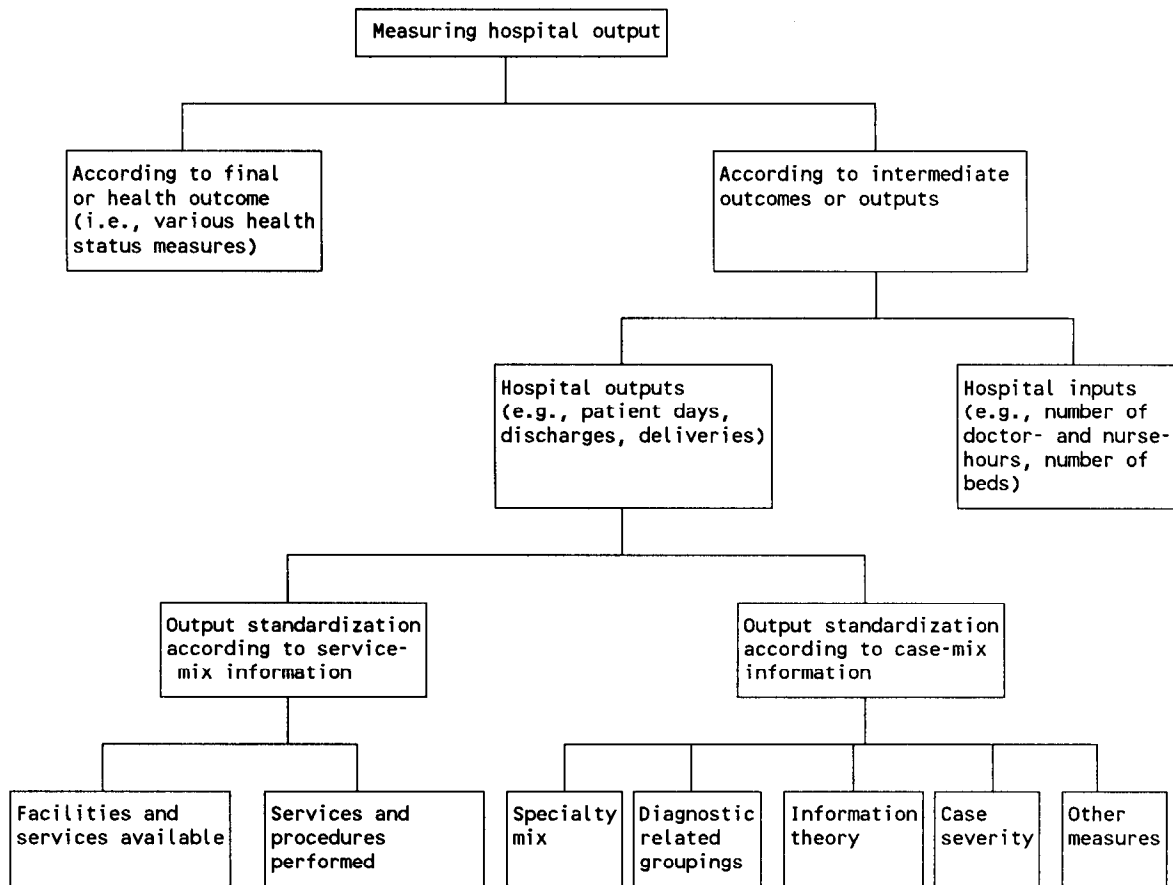
- (a) changing technology;
- (b) varying qualities of care;
- (c) varying case-mix
- (d) varying case complexity or case severity;
- (e) varying institutional characteristics—size, teaching status, location, composition, ownership, and so on. [p.872]

The author identifies two methods to standardize hospital output: service-mix and case-mix. The service-mix approach uses information about the types of services offered by the hospital (e.g., obstetrics and gynecology, oncology) as a basis for standardizing hospital output. Vitaliano (1987) calls this a supply-side method because it does not directly take into account utilization information, which is a function of demand as well as supply. Tatchell argues that the existence of common services among providers does not imply that the mix and complexity of cases, as well as the quality of care, are similar among providers. He also points out that knowing whether or not certain

services are offered by a particular provider is not sufficient to characterize its output, because there is a lack of information about the extent to which those services are actually used.

The case-mix approach attempts to standardize hospital output according to the mix of cases actually treated in hospitals. Vitaliano (1987) calls this method a demand-side one. In fact, both methods reflect demand and supply forces, since the services offered by hospitals generally respond to demand in the longer run. The case-mix approach, however, captures supply-demand interactions better because it is based on utilization information. Under the case-mix approach, Tatchell distinguishes several methods; they are: specialty mix, diagnostic related groupings, information theory, case severity, and other measures. The definitions, advantages, and disadvantages of these methods, as well as the studies that have used them, are reviewed in detail by Tatchell. Exhibit 8 summarizes the approaches available to standardize hospital output, according to Tatchell's review.

EXHIBIT 8
Approaches for Standardizing Hospital Output
According to Tatchell (1983)



The remainder of this section focuses on the use of the case-mix approach to control for variations in output when measuring health facility efficiency.

When comparing efficiency among providers, the problem of variations in case mix can be overcome in part through econometric estimation of multi-product cost functions (see discussion in Section 4.5). This approach has been used by Grannemann et al (1986), Vitaliano (1987), Eakin and Kniesner (1988), Eakin (1991), and many others. Because of data limitations, however, while it is sometimes possible to obtain data about composition of output, it is usually difficult to gather information on output complexity. Complementary techniques must then be used in studies of health facility efficiency. These are described below.

Register and Bruning (1987) suggest that case mix differences may introduce a bias in their comparative study of efficiency between for-profit and not-for-profit U.S. hospitals. Lacking information on case mix of individual hospitals, they distinguish among hospitals based on a procedure that uses case mix proxies to limit the confounding effects of case heterogeneity on efficiency measures. They assume that urban and rural hospitals differ in case mix and thus drop rural hospitals from their sample. Next, they eliminate from the sample all long-term and federal hospitals. Then they restrict their sample to hospitals which provide a specified set of services, thus eliminating those that are either too basic or too "high-tech".¹⁸ Finally, they posit that hospital size, as measured by the number of beds, also is associated with case mix. They therefore limit their sample to hospitals within the range of 100 to 250 beds.

In order to check their hypothesis that hospital size is associated with case mix, the authors use a rather questionable method, however. They group hospitals by size and compare the groups on the basis of bed-to-doctor and bed-to-nurse ratios. They interpret differences in ratios as differences in case mix, a rather questionable conclusion. Contrary to their intentions, this procedure may introduce a bias into their analysis, because beds-to-labor ratios may reflect not only case mix but also efficiency differences.

The authors then use a multiple-output, multiple-input approach for estimating hospital technical efficiency.¹⁹ This allows them to control for case composition, while the previous procedure (if the bed-to-labor argument is assumed to be appropriate) for restricting the sample allows them to control for complexity, as well.

Register and Bruning found that no significant difference in technical efficiency existed between for-profit and non-profit hospitals.

In their study of economic efficiency of U.S. hospitals, Eakin and Kniesner (1988) also confront the problem of case-mix differences among facilities:

¹⁸ Register and Bruning (1987), p.902.

¹⁹ The technique and results are described in section 4.6.

The output data indicate that hospitals are a heterogeneous group with respect to the mix of cases treated. Consequently, we adopt a multiple output specification in this study. In particular, we use four categories of outputs—general medicine, obstetrics and gynecology, weighted surgery, and outpatient visits. [p.587]

The above procedure allows the authors to control (very crudely) for case composition. In order to control for case severity and factors other than case mix influencing hospital costs, they proceed as follows:

Three additional variables—TEACH, a dummy variable indicating medical school affiliation, ALOH, the average overall length of hospitalization, and OR, the average overall occupancy rate—are included in the list of regressors to control for case severity and capacity utilization. [p.587]

The variable TEACH is used to capture the (positive) effect on hospital costs of teaching activities. As pointed out by Vitaliano (1987), 'The teaching-related costs of a medical school are likely to exert an upshift, "ceteris paribus", on hospital costs. Normally, the expected sign of the medical school variable would be positive' [p. 310]. The length of hospitalization variable is introduced by Eakin and Kniesner to control for severity; however, this variable may also represent differences in medical practice not attributable to severity but which may reflect differences in efficiency. The occupancy rate variable is introduced to discriminate between technical and economic efficiency, a useful distinction. A hospital can be technically efficient (i.e., it can operate on its production possibilities frontier) but, because of low occupancy, it may be economically inefficient. Because the demand for inpatient services is to a large extent exogenous to the facility, a particular hospital located in a low demand area may be able to attain maximum technical efficiency yet unable to raise its economic efficiency.²⁰

Vitaliano (1987) adopts a service-mix approach to control for case-mix and its effects on hospital efficiency which draws from Francisco (1970) and Tatchell (1983). For each hospital in his sample he constructs an unweighted index of 11 available hospital facilities or services and uses that index as an independent variable to estimate a total cost function. He finds that the service-mix variable is statistically significant and has an important positive effect on total cost.²¹

In sum, there are several methods available to control for case mix when evaluating health provider efficiency.²² They include:

- (1) Limiting the sample of providers to be compared on the basis of efficiency to those that provide a similar case mix;

²⁰ The findings by Eakin and Kniesner are discussed below, in Section 4.6.

²¹ Additional findings from Vitaliano (1987) are presented in Section 4.6.

²² For a comprehensive review of methods available, see Tatchell (1983).

- (2) In econometric studies of efficiency, including as explanatory variables indicators of case mix such as the types and volumes of outputs produced, measures of output complexity, and qualitative variables indicating whether the provider performs other activities that affect resource use, such as the provision of medical education; and
- (3) Limiting the comparison of efficiency to a few, well-defined medical services (e.g., appendectomies, hernia repairs, etc.), as was done by Pineault et al. (1985). When using the first method, case mix information should be derived based on the providers' outputs and not on the inputs (as was done in some of the studies described above), since input usage is an element of productive efficiency.

4.3 THE MEASUREMENT OF EFFICIENCY AND INPUT PRICE ISSUES

As was discussed in Section 3, knowledge of input prices is necessary to compare economic efficiency among providers. Further, to assess social economic efficiency, knowledge of shadow prices of inputs is needed, as well.

The DEA technique, explained in greater detail later in Section 4, limits itself to the calculation of technical efficiency, and therefore does not use any input price information. Ratio analysis, described below, can require input price data. While some examples were found in the literature where provider expenditure information was used for the computation of efficiency ratios, no examples were found where input prices were used explicitly in the analysis. Similarly, no discussion was found among ratio analysis studies about the issue of departures between market and social prices of inputs and the associated efficiency implications. The literature on efficiency measurement using econometric estimation of cost functions does address the input price issues discussed in Section 3. Examples of approaches are provided in Section 4.5 below.

4.4 MEASUREMENT OF EFFICIENCY: RATIO ANALYSIS

The measurement of productive efficiency through ratio analysis generally consists of computing and comparing one or both of the following two types of ratios: input to output ratios and cost of inputs to output ratios. An obvious and important limitation of this method is its inability to deal with multiple-output production. The first type of ratio—input to output—approximates technical efficiency; the second—cost of inputs to output—approximates economic efficiency.

There is a rich literature on the measurement of health services efficiency using ratio analysis. Barnum and Kutzin (1990) provide an excellent review of that literature for developing countries. To illustrate the types of analysis performed in ratio analysis, two papers about the U.S. are briefly described below.

Sear (1991) conducted a study of efficiency and profitability of investor-owned and not-for-profit hospitals using a sample of 142 hospitals obtained in 1988 by the State of Florida Health Care Cost Containment Board. He used three measures of efficiency: (1) the total number of full-time

equivalent (FTE) personnel per active bed; (2) the number of man-hours per adjusted patient day;²³ and (3) the total wages paid per adjusted patient day. He found that investor-owned hospitals used significantly fewer FTE staff per bed, had significantly fewer man-hours per adjusted patient day, and paid significantly less in wages (this last finding — a difference in the price of an input — does not say anything about relative economic efficiency).

Lee et al. (1984) conducted a comparative study of efficiency between free-standing ambulatory care and hospital-based care using data from three hospitals and three neighborhood health centers participating in the Bedford-Stuyvesant and Crown Heights Demonstration Project. Using a simple method for allocating fixed and indirect costs, the authors computed the average cost per visit in hospitals and health centers. They found wide variability in costs and no clear pattern emerged between the two types of facilities. Using other methods for apportioning various costs to visits, the authors demonstrated how sensitive the results were to the cost allocation rule employed.

Ratio analysis has several advantages, including its conceptual simplicity, ease of computation, low cost, and the fact that it can be performed on small samples.²⁴ Ratio analysis has several limitations, however. A first, and important one, is the method's inability to deal with multiple-output production, as noted above. Also, in some instances, ratio analysis requires that certain costs of multi-product providers be allocated among services, an exercise that usually requires arbitrary assumptions about allocation rules. Often, the results are very sensitive to the allocation rule used. Also, ratio analysis is not immune to the problem of quality and case mix variations among providers. Unless adequately done (two totally different kinds of ratios are involved), ratio analysis may also fail to distinguish between technical and economic efficiency.

Depending on the circumstances, the advantages of ratio analysis may prevail over its shortcomings and its use may thus be warranted. Additionally, ratio analysis is a powerful tool for assessing efficiency within a health care institution. A thorough discussion of the advantages and limitations of this technique can be found in Barnum and Kutzin (1990).

4.5 MEASUREMENT OF EFFICIENCY: ECONOMETRIC ANALYSIS OF HEALTH FACILITY COSTS

The statistical study of health facility costs, especially hospitals, has received a great and growing amount of attention in the U.S. literature over the past three decades. Cowing et al. (1983), citing Feldstein (1974), attribute this interest to two primary factors:

²³ Adjusted patient day was defined as follows:

$$AdjPatDay = \sum Patient\ Days \cdot \frac{1 + outpatient\ revenue}{1 + inpatient\ revenue}$$

Presumably, this ratio constitutes an attempt (however questionable) to capture in a single measure (AdjPatPay) the volume of both outpatient and inpatient output.

²⁴ If standard statistical confidence is sought for differences in ratios, however, the sample of providers must be large enough.

...the emergence of health care as a major national public policy concern, and the use of the health sector to test new economic models dealing with such "nonstandard market" characteristics as non-profit institutions, non-price rationing, third-party financing of services, and inherent uncertainty on the part of the consumer. [p. 257]

There are a number of excellent surveys of the hospital cost literature, including Cowing et al. (1983), Feldstein (1974), and many others. The interested reader is referred to those reviews to obtain a comprehensive picture of the work done prior to 1983.

Many additional studies have been conducted since the Cowing et al. (1983) review, and with an increasing degree of sophistication. Providing a comprehensive survey of those, of course, falls outside of the scope of this paper. Instead, this section reviews a small number of recent studies of health facility cost which use somewhat different methodologies and which are considered particularly relevant for the study of health services efficiency in the developing world. They include Grannemann et al (1986), Vitaliano (1987), Eakin and Kniesner (1988), Frank and Taube (1987), and Eakin (1991).

Grannemann et al. (1986), in their statistical study of costs and efficiency of a sample of U.S. hospitals, introduce a "modified translog" total cost function where outputs enter directly, and not in logarithmic form, to permit zero output levels. Capital is included as an independent variable, reflecting their assumption that capital is not fully exogenous to the hospital. This total cost function is therefore a long-run function which allows for endogenous variations in a hospital's capital stock. They also include various measures of hospital output, including emergency care, home visits, and two measures of inpatient care, to account for differences among hospitals in average length of stay.²⁵

Contrary to Eakin and Kniesner (1988), Grannemann et al. had poor data on input prices. This led them to exclude from their cost function interaction terms between input prices and output levels. This introduces an important theoretical constraint in their work.²⁶

²⁵ A compact expression for the logarithm of the total cost function used by Grannemann et al is

$$\ln C = \ln A + \sum_i \alpha_i \ln P_i + F(Y, D, CM, R, Z) + \epsilon$$

where A contains a vector of various factors that are assumed to affect the level of costs, but not the shape of the cost function with respect to outputs; P_i is the price of the i^{th} production input; Y is a vector of hospital outputs (Y_1 - Y_3 are inpatient days, by type, Y_4 is outpatient visits, and Y_5 is emergency department visits); D is a vector of inpatient discharges by type, CM is a matrix of case mix variables, R_1 and R_2 are vectors of variables indicating hospital's sources of revenue; Z is a vector of miscellaneous other outputs produced in some hospitals; and ϵ is a random error. The F function includes single, squared, cubic, and interaction terms for the Ys, single and squared terms for the Ds and the Ys and selected interaction terms between: the Ds and the CMs, the Ys and the CMs, and the Ys and the Rs.

²⁶ The authors describe these constraints as follows:

...hence, the production function associated with this cost function by duality is homothetic. This implies that the cost minimizing mix of inputs is unaffected by the volumes or mix of outputs produced. It also means that changes in input prices will affect cost estimates only by a scale factor; they will not change the relationship between marginal and average incremental cost or measures of economies of scale.

The authors estimate various measures of hospital performance, including marginal cost, average incremental cost, and product-specific economies of scale and scope (see definitions of terms in Section 3). Estimation of the cost function and associated measures is done using data from the 1981 American Hospital Association's *Annual Survey*. Estimation is done using ordinary least squares.²⁷

Grannemann et al. find that case mix measures of both inpatient and outpatient care are important, and that characterizing outpatient care volume according to both patient days and discharges helps explain cost variations among hospitals. They also find that estimates of marginal costs for these two inpatient quantity measures are inconsistent with a reimbursement system which pays on a discharge basis only while reimbursing nothing at the margin for additional days of care. The results suggest economies of scale (declining average incremental cost) for emergency outpatient care but no economies of scale for other kinds of ambulatory care. Diseconomies of scope are found between inpatient and outpatient activities.

Vitaliano (1987) argues that econometric studies of efficiency which use a multiple output approach are plagued by the statistical problem of multicollinearity among outputs, which hinders the interpretation of regression coefficients. He also points out that hospital cost behavior should be studied in the context of a system of equations, where price, cost, and output are determined jointly; a single equation is a reduced form of the system, a fact that hampers the interpretation of statistical coefficients.

Vitaliano uses a sample of 166 New York State hospitals. He argues that, because prices are exogenous to the hospital (due to prospective reimbursement) and hospital "output is non-storable and supplied on demand", estimation of a single cost equation is appropriate.²⁸ He thus proposes a minimum total cost equation with a single output, to conduct a comparative study of economic efficiency. Next, apparently owing to data problems, he proposes to use number of available hospital beds as a proxy for hospital output. Beds, a capacity measure, is more weakly related to costs than output is. Other explanatory variables included as regressors are: the teaching condition of the hospital (MEDSCHOOL); a case mix index of services provided in the hospital (FACILITIES); a regional dummy variable (URBAN); and a measure of the hospital's monopoly and monopsony power (SHARE), as measured by the hospital's share of the total supply of beds in

²⁷ As is acknowledged by the authors, the possible endogeneity of output may bias the regression coefficient on the output terms in a single-equation OLS model. They argue, however, that this bias may be small because third-party reimbursement weakens the relationship between the amount reimbursed (and thus cost) and quantity demanded.

²⁸ Vitaliano also supposes that hospitals are cost minimizing, an arbitrary assumption which may hamper the interpretation of statistical results if wrong. He argues 'Although almost all the hospitals included in this study are not-for-profit, the regulators allow them to keep any "profit" and impose financial penalties for cost overruns. Thus it is not unreasonable to assume least-cost behavior on the part of the hospitals.'

its county.²⁹ Vitaliano argues that a monopsony or oligopsony hospital may be able to exert downward pressure on factor prices. His estimation technique is weighted least squares.

With his quadratic cost function, the author obtains a U-shaped average cost function. However, when he uses a different specification for the cost function (log form), he finds a declining average cost curve and, thus, economies of scale. These results are not entirely surprising, since a quadratic function is likely to yield a U-shape average cost curve, whereas a logarithmic specification is not. He attributes the economies of scale to the presence of high fixed costs such as specialized personnel and equipment.

Eakin and Kniesner (1988) propose a general method for estimating hospital cost functions which allows for the possibility that hospitals may not be cost minimizers. They point out that cost minimization is a special case of a more general case where cost minimization does not necessarily occur. They point out that "erroneously assuming cost minimization leads to inaccurate estimates of factor substitution possibilities." [p.584]

The authors list several reasons why hospitals may minimize neither their own private costs nor the social cost of production. Failure to minimize private costs may result from: differences between expected and actual input prices; systematic biases in hiring resulting from satisficing behavior on the part of managers; and mismeasurement of prices. Failure to minimize social costs may result from any of the above reasons, as well as from systematic over- or under-valuation by the firm of the social price of resources.

Given a technology and input prices, costs are minimized when the ratio between any pair of input prices equals the ratio between the inputs' marginal rate of technical substitution (see Section 4.6),

$$\frac{w_i}{w_j} = \frac{F_{Xi}}{F_{Xj}}$$

where w is input price and F_X denotes the marginal product of factor X .

²⁹ His final total cost equation is a quadratic functional form:

$$C = a + b \cdot BEDS + c \cdot BEDS^2 + d \cdot MEDSCHOOL + e \cdot FACILITIES + f \cdot URBAN + g \cdot SHARE + \epsilon$$

where ϵ is a random error term and the other variables are as defined in the text.

Misperception of the shadow price of inputs by hospital managers leads them to equate the rate of technical substitution to the ratio of their perceived input prices:

$$\frac{w_i^{sh}}{w_j^{sh}} = \frac{w_i + \theta_i}{w_j + \theta_j} = \frac{F_{Xi}}{F_{Xj}}$$

where the superscript sh denotes the firm's shadow price and θ is the "shadow price divergence parameter." [p.584]

Eakin and Kniesner establish a relationship between the hospital's observed cost function, with observed total cost and observed input prices, and the shadow cost function, with shadow cost and shadow prices. This relationship reduces to the traditional minimum cost specification when $\theta_i = 0$, i.e., when the hospital's shadow price and the actual price of inputs coincide. They estimate a system of equations which include the observed cost function and three observed factor share equations. The total cost function is specified as a "hybrid" translog where, to accommodate zero output levels, they enter the actual values of the outputs instead of their logarithm.³⁰ The system is estimated via the method of seemingly unrelated non-linear regression.

Estimation of the non-minimum cost function requires data on observed costs, observed input cost shares, input prices, and output levels. In their specification they include four outputs (general medicine, obstetrics and gynecology, weighted surgery, and outpatient visits) and four production inputs (non-physician labor, capital, physicians, and material). Shadow costs are, of course, not observed but estimated. Their sample consists of 331 U.S. short-term hospitals. Efficiency-related measures derived by the authors include estimates of the Allen elasticities of factor substitution, the price elasticities of factor demand, marginal costs, output cost elasticities, and overall economies of scale.

The authors measure allocative efficiency (AI) as

³⁰ The estimated system of equations includes the logarithm of the observed total cost C^{obs} equation and three observed factor share equations M_i^{obs} (four cost shares are observed -- i=labor, capital, materials, and physicians-- but only three equations are independent and thus included in the system (the physicians equation is excluded)). The equation for the logarithm of observed total cost is

$$\ln C^{obs} = \ln C^{sh} + \ln \left[\sum_i^n (w_i/w_i^{sh}) \cdot \beta_i + \sum_j^n \beta_{ij} \cdot \ln(w_j + \theta_j) + \sum_k^m \delta_{ik} \cdot Y_k \cdot \ln(w_i + \theta_i) \right]$$

where Y_k denotes the level of the k^{th} output and w_i^{sh} is the hospital's shadow price of the i^{th} input. The factor share equations included in the system are

$$M_i^{obs} = M_i^{sh} \cdot \left[\sum_h^n (w_h/w_h^{sh}) \cdot \beta_h + \sum_j^n \beta_{hj} \cdot \ln(w_j + \theta_j) + \sum_k^m \delta_{hk} \cdot Y_k \cdot \ln(w_h + \theta_h) \right]$$

$$AI = \frac{(\hat{C}^{obs} - \hat{C}^{min})}{\hat{C}^{min}}$$

where \hat{C} denotes estimated total cost.

They find empirically that hospitals undervalue the cost of capital and overvalue the price of physicians. Hospitals thus tend to over-employ capital and under-employ physicians. Thus, the hospitals in the sample do not minimize cost relative to market (or observed) prices. This results in an allocative inefficiency which is equal to about five percent of the total observed cost. Eakin and Kniesner also find that input concepts, such as elasticities of substitution and factor demands, are sensitive to model specification. However, they conclude that:

A researcher who is primarily interested in output concepts, such as marginal cost and scale and scope economies, does not need to use a non-minimum cost function, even if cost minimization cannot be readily assumed [p.597].

In the case of developing countries, however, where markets are less perfect than in the industrialized world, gaps between social, market, and perceived input prices may be large, particularly among governmental providers. Wrongly assuming cost minimization, and thus using a standard minimum cost function, could result in large errors in both input and output measures. The non-minimum cost approach adopted by Eakin and Kniesner therefore may be warranted for estimation of productive efficiency in developing-country health services.

Frank and Taube (1987) study technical and allocative efficiency in the production of outpatient mental health clinic services, using data from 755 clinics obtained through a 1982 survey by the U.S. National Institute of Mental Health. Their study focuses on the estimation of production functions for mental health visits. Like Eakin and Kniesner (1988), they are interested in exploring provider departures from cost minimizing behavior. They also do so by comparing the ratio between factor prices with the ratio between the factors' marginal products. As explained above, cost minimization is present when those two ratios are the same for all pairs of variable inputs. Differences in price-marginal product ratios mean that certain production inputs are either over- or underemployed. Frank and Taube are also interested in system efficiency (see discussion in Section 2), and thus study the extent of economies or diseconomies of scale in production of mental health visits. Finally, they study the determinants of efficiency by focusing on differences in productivity between government-run county and private clinics.

The authors use two different empirical specifications for the production functions of the clinics: a Cobb-Douglas and a transcendental production function.³¹ Estimation of the two alternative production functions is done using ordinary least squares.³² They find decreasing returns to scale with both specifications, and greater productivity of private clinics, as evidenced by an ownership dummy included in the production functions. Finally, they compare input price ratios with marginal product ratios (evaluated at mean values) and find that they differ for the case of physician and other clinical staff, physicians being over-employed. This signals a departure from cost minimizing behavior. With regard to their empirical specifications, they find that the simpler Cobb-Douglas model performs better than the alternative specification, as measured by the goodness of fit criterion.

Eakin (1991) uses the results from Eakin and Kniesner (1988) to study the determinants of economic efficiency in hospitals. He does this by regressing their estimated measure of allocative inefficiency (see above) against several hospital and market characteristics predicted by economic theory to be determinants of efficiency. The independent variables include ownership (i.e., a for-profit dummy and a dummy indicating whether or not the hospital is operated by a religious entity; regulatory factors (e.g., the presence of a Certificate of Need program which controls hospital capital outlays); competitive factors, as measured by the percentage of the population enrolled in HMOs; factors characterizing the sources of hospital revenue, including the proportion of patients who self-pay, who are privately insured, and who are MEDICARE and MEDICAID beneficiaries; measures of hospital size (number of beds); and regional dummy variables. Estimation is done using estimated generalized least squares (EGLS).

4.6 MEASUREMENT OF TECHNICAL EFFICIENCY: DATA ENVELOPMENT ANALYSIS AND RELATED TECHNIQUES

This section presents the work of various authors in estimating technical efficiency, using a family of operations research techniques called Data Envelopment Analysis (DEA). Unlike econometric techniques for the estimation of production or cost functions, this linear programming technique permits the study of production efficiency without the need to make any assumptions

³¹ The Cobb-Douglas production function is

$$\ln Q = \ln A + \alpha \cdot \ln X_1 + B \cdot \ln X_2$$

The ratio of the marginal products is $(\alpha X_2)/(B X_1)$ and the elasticity of scale (ES) is $\alpha + B$. The transcendental production function is specified as

$$\ln Q = A + \alpha \cdot \ln X_1 + B_1 \cdot X_2 + B_2 \cdot X_2^2$$

The marginal product ratio is $\alpha/(2B_1B_2X_1X_2)$ and $ES = \alpha + (B_1 + 2B_2X_2)X_2$. If $ES < 1$ then there are decreasing returns to scale, while $ES > 1$ implies increasing returns to scale.

³² The use of OLS has been questioned in this context. The authors indicate that this practice may introduce a bias due to the joint determination of inputs and outputs. Citing Reinhardt (1972), however, they point out that the bias is small if factor and product prices vary substantially across the sample observations, which happens to be the case in their sample of mental health clinics.

about the technology of production or the presence of cost-minimizing behavior. DEA is a non-parametric technique, that is, one which estimates efficiency without the need to estimate any parameters, such as those of production or cost functions. Finally, and also contrary to econometric techniques, DEA is a deterministic technique, which, as such, does not include explicitly a statistical error term reflecting measurement or sampling error.

The papers reviewed in this section include Sherman (1984), Register and Bruning (1987), Valdmanis (1990), and Byrnes and Valdmanis (1989). A review of studies that use DEA to measure technical efficiency in health care organizations is provided by Rosko (1990). Before proceeding with a review of the above references, however, a conceptual discussion about the DEA method is provided.

Farrell (1957) developed a framework, susceptible to empirical estimation, to distinguish between technical and economic efficiency.

His approach can best be explained graphically, by considering a process that produces a single output, Q , with two inputs, X_a and X_b (see Exhibit 9). As explained in Section 3, all points along the isoquant, such as A and B, are technically efficient but only one point (point A, at the intersection of the isoquant and isocost lines) on the isoquant is economically efficient. Consider production of quantity Q at point F. At F there is technical *and* economic inefficiency. Farrell measures technical efficiency as the ratio OB/OF . Because at F more inputs are used to produce Q than is technically necessary, OF is greater than OB , OB/OF is less than one, and thus F is technically inefficient. But at point B (as well as at point F, of course) there is also economic inefficiency, because

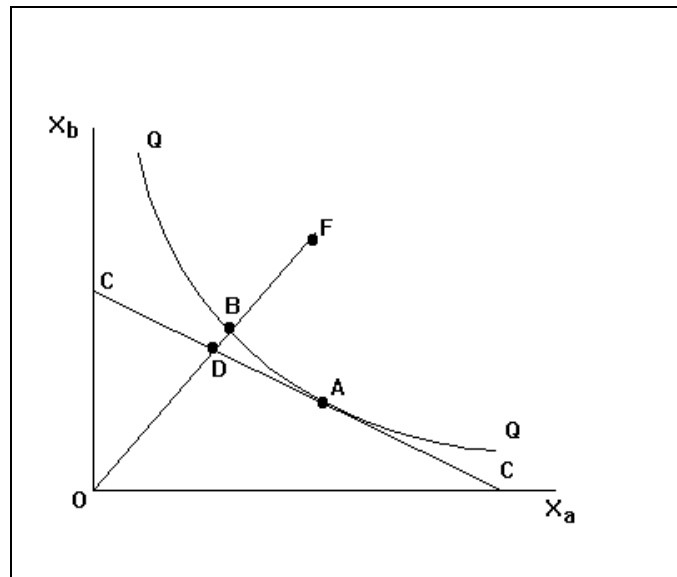


Exhibit 9: A graphical interpretation of technical and economic inefficiency.

the least costly combination of inputs to produce Q is not chosen. Economic efficiency of point B is measured as OD/OB , where D is a point on the isocost line. Since OB is greater than OD , OD/OB is less than one, and therefore there is economic inefficiency. Farrell thus states that a provider's deviation from minimum cost (or from maximum economic efficiency) can be attributed to, or decomposed into, technical and economic inefficiencies. Thus:

$$OE = TE \cdot AE$$

or overall economic efficiency (OE) equals technical efficiency (TE) times economic efficiency (AE). Using the example from the figure, overall efficiency of production at point F would be

$$OE_F = \frac{OD}{OF} = \frac{OB}{OF} \cdot \frac{OD}{OB}$$

The Implementation of DEA. Consider a set of several providers (say, eight of them) each producing different quantities of a single, homogeneous output, and consuming different combinations of the two production inputs, X_a and X_b , as shown in Exhibit 10. How efficient are these providers? Let us consider *technical efficiency* (Farrell discards the usefulness of estimating economic efficiency, stating that "[economic] efficiency is a measure that is both unstable and dubious of interpretation.(p.261)" His approach, as well as subsequent approaches and the DEA technique, are concerned only with technical but not with economic efficiency.) Because the quantities of output Q differ among the providers, it is

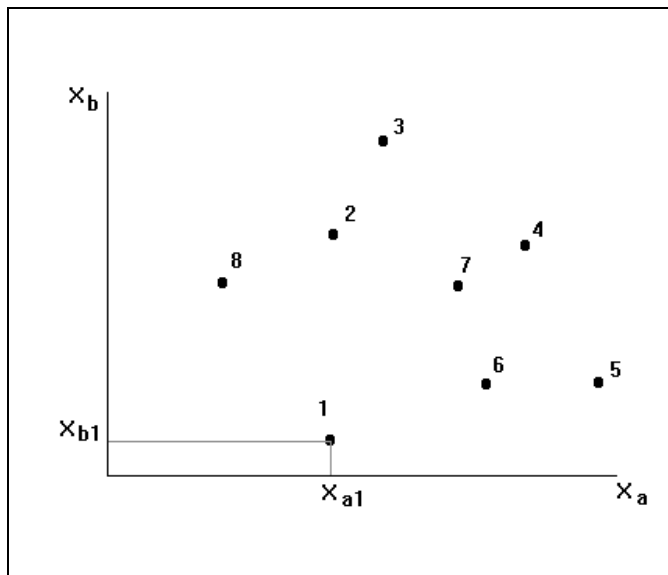


Exhibit 10: Several Providers Producing Different Quantities of Single, Homogeneous Output (Q) With Two Inputs (X_a and X_b)

not possible, at least graphically, to determine the relative technical efficiency of each: differences in input use may be due to either differences in the volume of output or differences in the technical efficiency of input use, or both. To facilitate comparison, the providers can be *normalized* according to their output level by dividing their consumption of inputs by their respective output. For example, for provider 1 we do X_{a1}/Q_1 and X_{b1}/Q_1 , which yields point P_1 in Exhibit 11.³³ All other points in Exhibit 11 are similarly obtained. Now, the technical efficiency of these providers can be gauged more easily.

At this point, it is important to distinguish between absolute and relative technical efficiency. Ideally, one would like to determine how each of the eight providers fares relative to the theoretical, technically most efficient production frontier. Unfortunately, as Farrell points out, it is often difficult to know what the theoretically efficient production practice is, and this difficulty increases with the level of complexity of the production process (as happens when the number and diversity of production inputs goes up, a phenomenon rather typical in the production of health care). Because of this difficulty, Farrell discards the possibility of devoting effort to the determination of the theoretically efficient isoquant. In any case, whether that isoquant is known or not, it would definitely fall to the South-West, or at the South-Western edge, of all observed points, as depicted in Exhibit 11.

³³ Exhibit 11 and the numerical examples provided below are adapted from Charnes et al. (1978).

Departures from the theoretically efficient isoquant being impossible to compute, Farrell proposes to focus on the analysis of technical efficiency of production units *relative to the most efficient observed practice*. The solid line of Exhibit 11 constitutes precisely that. It is obtained simply by joining with straight segments the south-westernmost points in the figure. The line thus obtained is the best practice, technically efficient unit isoquant. Thus, points P_3 , P_4 , P_5 , and P_6 all have a technical efficiency of 100% relative to the eight providers in the set. In contrast, P_1 , P_2 , P_7 , and P_8 all are technically inefficient *relative to the providers on the efficient practice unit isoquant*.

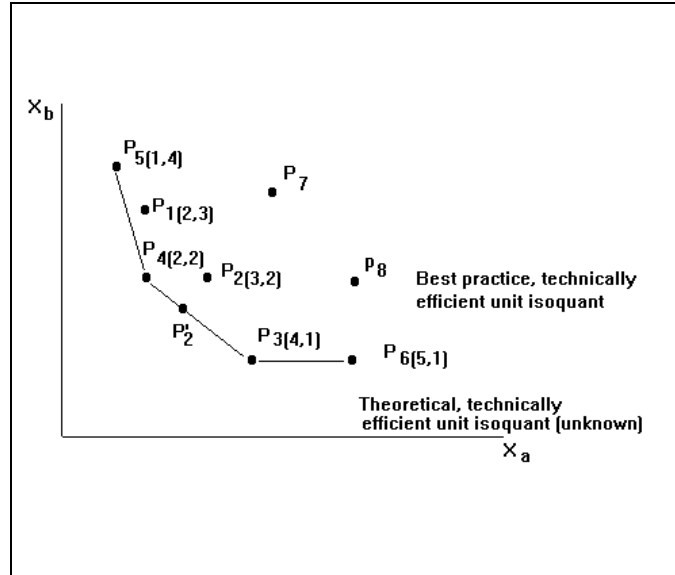


Exhibit 11: Efficiency Points and Unit Isoquant (adapted from Charnes et al., 1978)

Using linear programming techniques, such as any *adjacent extreme point method* like the *simplex* or *dual* methods, it is possible to compute an index of technical efficiency for each provider in a set. For example, consider provider 2 (P_2). Using such methods, provider 2's position in the input plane can be expressed as

$$P_2 = \frac{5}{6} \cdot P_4 + \frac{1}{3} \cdot P_3$$

That is, it is possible to express P_2 in terms of an optimal basis (adjacent points P_3 and P_4 , both of which are 100% technically efficient). The sum of the two weights ($5/6$ and $1/3$) is $7/6$. By dividing the above expression by $7/6$, one obtains

$$\frac{5}{7} \cdot P_4 + \frac{2}{7} \cdot P_3 = \frac{6}{7} \cdot P_2 = P_2'$$

The above result indicates that if P_2 had been producing as efficiently as P_3 and P_4 , then it should have been able to produce its output with only $6/7$ of the amount of inputs X_a and X_b it actually used. The index of technical efficiency (ITE) of P_2 is therefore $6/7$, or about 86%. Notice that $6/7$ is equivalent to the technical efficiency ratio proposed by Farrell and described in Exhibit 9, OB/OF. A similar calculation can be performed to show that P_1 can be expressed as a linear combination of the technically efficient basis P_4 and P_5 , yielding, coincidentally, an ITE of 86%. Using the linear programming methods described above, the ITE of P_7 and P_8 could similarly be computed.

Building on the work of Farrell (1957), Charnes et al. (1978) proposed a formal linear programming approach for studying technical efficiency of "decision making units", or "DMUs", allowing for multiple output production (a common characteristic of most health providers). They proposed an iterative algorithm where, in each iteration, the technical efficiency of a DMU is computed relative to all the other DMUs in the set. The efficiency measure is obtained as "the maximum of a ratio of weighted outputs to weighted inputs subject to the condition that the similar ratios for every DMU be less than or equal to unity (p.430)".³⁴

According to the Farrell (1957) and Charnes et al. (1978) approach, a facility with an ITE of one reflects the best technical practice within the set studied. It would, of course, be possible to find facilities outside of the set with even greater technical efficiency. In the case of those facilities in the set with ITE<1, DEA unambiguously indicates that they can improve their technical efficiency by reducing the quantities of one or more inputs while keeping output levels constant.

DEA has several advantages. It permits estimation of technical efficiency in the case of multiple output health facilities. Because of that, the analyst can control for case mix by specifying in the output vector detailed information about the types and complexity of cases treated. DEA requires no assumptions about the technology of production or about whether the providers seek to maximize profits (or minimize costs).

³⁴ The precise formulation of the Charnes et al. (1978) approach is

$$\max h_0 = \frac{\sum_{r=1}^s u_r \cdot y_{r0}}{\sum_{i=1}^m v_i \cdot x_{i0}}$$

subject to:

$$\max h_0 = \frac{\sum_{r=1}^s u_r \cdot y_{rj}}{\sum_{i=1}^m v_i \cdot x_{ij}} \leq 1;$$

$$\text{for } j = 1, \dots, n; u_r, v_i \geq 0; r = 1, \dots, s; i = 1, \dots, m.$$

where the subindex 0 is used to denote the facility for which technical efficiency is being calculated and where "the y_{ij} , x_{ij} (all positive) are the [observed] outputs and inputs of the j^{th} DMU and the $u_r, v_i \geq 0$ are the variable weights to be determined by the solution to this problem --e.g., by the data on all of the DMUs which are being used as a reference set. The efficiency of one member of this reference set of $j=1, \dots, n$ DMUs is to be rated relative to the others" (p.430). According to this approach, the solution to the technical efficiency calculation problem is obtained by solving, in any order, n linear programming, constrained maximization problems.

It is important to understand, however, that no firm conclusions can be inferred about the economic efficiency of the facilities based on their DEA-computed index of technical efficiency, especially if their choice of production inputs places them far apart along the production possibilities frontier.³⁵ This has been illustrated in Exhibit 7 of Section 3. There, providers 2 and 3 have equal technical efficiency which is higher than provider 1's. Because production input X_a is sufficiently inexpensive relative to X_b , however, provider 2—and provider 1—are economically more efficient than 3. DEA would assign 2 and 3 equal technical efficiency and would rate 1 as inferior. Economically, however, 1 would be superior to 3 (but not to 2).

Sherman (1984) introduces DEA analysis to the estimation of hospital technical efficiency. He uses a sample of six Massachusetts teaching hospitals and defines their outputs and inputs for the medical-surgical area through discussions with a panel of experts. Using DEA, two out of the six hospitals are found to have relative technical efficiency ratings of less than one. Sherman compares these results with the ranking used by the Massachusetts Rate Setting Commission to measure hospital efficiency. He demonstrates (not surprisingly) that the Commission's ratios (average cost per patient and per patient day) fail to distinguish technical inefficiencies. Because his analysis is limited to the estimation of technical efficiency, Sherman is unable to judge the validity of the Commission's ratios from an economic efficiency standpoint.

Sherman uses his results to urge the managers of the relatively inefficient facilities to examine their efficiency. Moreover, the results can be used to tell the managers which facilities' input proportions they ought to copy. He also uses the findings to re-examine technical efficiency following actual or potential management measures. He thus illustrates how DEA can be used to simulate the technical efficiency implications of various management practices such as reducing the number of beds, or purchasing fewer selected supplies. However, because DEA does not take factor price information into account, recommendations aimed at improving technical efficiency (factor substitution) might in fact result in reduced economic efficiency.

Register and Bruning (1987) use the Farrell technique and DEA to compare technical efficiency between for-profit and non-profit U.S. hospitals, using data from a 1984 American Hospital Association Survey. The authors explain the maximization problem as follows:

...given all the hospitals included in the sample, and considering the amount of each input employed by each hospital producing one unit of output, what is the maximum output that can be produced? In short, can the system of hospital care production be improved by reallocating productive inputs from a given hospital to other hospitals? If so, the hospital under consideration is inefficient relative to others in the sample.
[p. 907]

From the total sample, the authors extract a subset of 457 hospitals (300 are non-profit) which they consider to be relatively homogeneous in terms of case mix. They proceed in two steps:

³⁵ Variations in input prices across providers would also preclude conclusions about relative economic efficiency based on technical efficiency measures.

first, they compute an index of technical efficiency (ITE) for each of the 457 hospitals of their sample. Second, to attempt to explain technical efficiency, they use ordinary least squares (OLS) to regress the dependent variable of ITE against a series of independent variables believed to affect technical efficiency. These variables include a dummy variable for hospital ownership; a measure of market concentration; hospital size, as measured by the number of beds; and interaction terms between these variables.

The innovative feature of the Register and Bruning study is that it not only computes technical efficiency but it also attempts to find the determinants of technical efficiency. One problem with this study, however, is that, like other studies of technical efficiency, it does not use any information about input prices and thus provides few clues about the providers' relative economic efficiency. Another shortcoming of the study (and the method) is that it uses a single measure of output, the number of bed days. Even if attempts are made to limit the sample to a relatively homogeneous set of facilities, the use of a single output measure may be a serious limitation to study efficiency of multi-product firms, such as hospitals, with a very heterogeneous set of outputs.

Valdmanis (1990) also uses DEA to measure technical efficiency of hospitals in relationship to ownership, using a sample of Michigan State hospitals in Standard Metropolitan Statistical Areas. This sample came from the 1982 American Hospital Association Survey of Hospitals. In contrast with Register and Bruning (1987), and like Sherman (1984), she uses a multiple-output, multiple-input version of DEA. Technical efficiency indices are computed for public and non-profit hospitals separately. She then uses a non-parametric test (Mann-Whitney) to assess the statistical significance of the differences between the efficiency indices obtained for the two hospital groups. Valdmanis also studies the relationship between the technical efficiency indices and the cost per adjusted patient day for the two groups. She finds a statistically significant, negative relationship between technical efficiency and cost per patient day.

Byrnes and Valdmanis (1989) use DEA to study technical efficiency for a group of 123 non-profit, non-teaching, California community hospitals. They also use linear programming and information about input prices and technology to compute economic efficiency for individual hospitals in the sample. Their measures of output include acute medical surgical discharges, surgical intensive care unit discharges, and maternity discharges. The authors distinguish between fixed and variable production inputs. Their analysis thus presupposes the presence of short-run input constraints which preclude the achievement of long-run minimum cost in these facilities. The variable inputs they consider include: registered nurse hours; management and administrative personnel hours; technical services personnel hours; aide and orderly hours; and licensed practical nurse hours. The fixed inputs chosen are the number of staff physicians and physicians with admitting privileges and the average staffed beds.

In order to compute overall economic efficiency for the sample, the authors first solve the linear programming problem of minimizing total variable costs given input prices and various constraints. Second, using the computed minimum cost, they obtain the overall economic efficiency index (OE) by dividing observed variable cost by minimum estimated variable cost. Next, they

compute technical efficiency (TE) with the DEA technique. Finally, they derive an allocative efficiency (AE) index by dividing OE by TE.

In sum, DEA is a useful, and rather innovative method for estimating relative technical efficiency of a group of medical providers. It is designed for processes with multiple outputs and inputs and permits a detailed specification of the case mix. It can also be used as a management tool to simulate technical efficiency implications of alternative management practices.

4.7 CONCLUSIONS

The review of the industrialized-country studies of health services efficiency has emphasized the study of hospitals. That is not surprising, considering that hospitals account for the lion's share of health expenditures in developed countries. Because the same is true in developing nations, the above literature review remains highly relevant for our purposes.

Similarly, a great deal of the developed-country literature, particularly in the U.S., has dealt with the comparative study of hospital efficiency between non-profit and for-profit facilities, or between government and private institutions. This focus is indeed relevant in the developing-country context and fits well within the aims of this review, i.e., to identify and use new methods to assess efficiency of government and private providers of care.

There are limitations, however, to the benefits that the developing-country research on this topic can draw from the research done in the industrialized world. In particular, many of the studies and techniques reviewed rely heavily on large and detailed data sources. For example, several studies done in the U.S. have benefitted from the presence of large, existing data sets collected routinely by various hospital and medical associations. Despite the presence of these data bases, however, several developed-country researchers still find themselves constrained by data limitations (e.g., Grannemann et al., (1986) could not obtain accurate data on input prices).

Poor data availability in developing countries is likely to increase the cost, while limiting the sophistication and predictive power of the studies that can be conducted there. Although utilization statistics can often be obtained at the central or facility level, information on input prices and, more generally, costs, is seldom available in routinely kept records. This implies that studies of health facility efficiency in developing countries will generally have an important data collection component, a factor that will heavily increase the overall costs of research. Further, because government facilities are generally subsidized from the central level, data collection efforts often have to combine facility data with information obtained from the central level.

The methodological problems described in this section—qualitative, case mix, and input price variations among providers—are certainly present in the developing country context. The lessons learned from the above review are certainly applicable to the case of poorer nations. By the same token, the techniques used in developed countries, with the caveat of data availability mentioned above, can also be used in developing countries.

The following section presents a review of studies of health facility efficiency in developing countries.

5.0 REVIEW OF SELECTED STUDIES ON HEALTH SERVICES EFFICIENCY FROM THE DEVELOPING WORLD

This section presents a review of selected studies of health provider efficiency in the developing world. The studies reviewed here are Rodríguez and Jiménez (1985), Anderson (1980), Dor (1987), Bitran and Dunlop (1989), Wouters (1990), Lewis et al. (1990), and Shepard et al. (1991). Barnum and Kutzin (1990) provide an excellent review of some of these and other studies, including Barnum (1990 and 1991). The section concludes with a summary of the review, from which conclusions are drawn for the phases two and three work.

5.1 LITERATURE REVIEW

Rodríguez and Jiménez (1985) conducted a comparative study of productivity between private and decentralized public Chilean hospitals. Their measure of productivity was the inverse of the average length of stay (LOS), shorter LOS being associated with higher productivity. The authors broke LOS into three components, according to the stages patients follow in a hospital: diagnosis (D), medical treatment (T), and recovery (R), such that

$$LOS = D+T+R$$

They posited that while T must be done within the hospital, D and R can partially be accomplished on an outpatient basis. For a given patient case mix, Rodríguez and Jiménez hypothesized that a series of individual specific variables, such as age and income, can influence D and R. They assumed that a patient's income is a close proxy for the type of health insurance or coverage he or she has. Thus income was assumed to be closely correlated with the out-of-pocket price of the services and therefore to influence D and R.³⁶ Illness severity and the types and amounts of medical inputs provided to the patients were also assumed to influence D, T, and R.

Using a total sample of five hospitals (three private and two public) and 369 patients, the authors used ordinary least squares (OLS) to regress length of stay on patient age and income, medical inputs (use of pharmacy and laboratory exams), and facility-specific dummy variables. Three separate equations were estimated for obstetrics and gynecology, surgery, and internal medicine.

The authors found that, other things in their model being equal, private hospitals had lower LOS and thus were more productive than public institutions. They cautioned, however, that their analysis was weakened (rather severely, most likely) by their inability to obtain accurate information about diagnosis, case mix, and severity of the health problem.

³⁶ According to this reasoning, one should also expect to see that T is affected by the financing modality, given the behavioral incentives to providers and clients associated with various forms of payment.

Anderson (1980) conducted a statistical study of hospital costs using a sample of 51 Kenyan hospitals with data from the years 1975-76. Using OLS, he estimated an average cost function where the dependent variable was average cost per patient day (ACPD) and the explanatory variables were the capacity (SCALE) (as measured alternatively by available and used beds)³⁷; the occupancy rate (OCR); the average length of stay (ALS); the number of outpatient visits per inpatient day (TOPPD); the number of satellite ambulatory facilities operating under the hospital (SAT); and the nature of the hospital (PHD) (provincial or non-provincial).³⁸

Anderson estimated four alternative specifications of the model, using alternative dependent and independent variables. He found that the hospitals were operating with increasing returns to scale, as evidenced by negative (decreasing average cost) and statistically significant coefficients associated with the scale variables. He also found that higher occupancy levels resulted in lower average costs, concluding that greater demand should be accommodated within existing hospitals rather than through new ones.³⁹ Outpatient activity was found to increase average cost, as expected. In contrast, length of stay did not come out statistically significant. Because the costs of a hospital and its satellite facilities were intertwined, the average cost used as a dependent variable included also the cost of the satellite facilities. The regression results showed that a greater level of satellite activity had a positive impact on aggregate average cost. Finally, provincial hospitals were found to have higher average costs than district and sub-district facilities.

Some rather dramatic limitations of the analysis, as pointed out by the author, included the inability to control for differences in quality of care, case mix, and input prices. Some or all of these shortcomings, however, also arise in the other studies described below.

Dor (1987) estimated average cost functions for a sample of 19 Ministry of Health (MOH) and Social Security (IPSS) urban hospitals in Peru. A key explanatory variable was patient flow (F), which is equal to the ratio between the number of hospital admissions (ADMISS) and the number of beds (BEDS). Other dependent variables included case-mix variables (share of admissions that were deliveries; share of admissions that were surgery), an affiliation dummy (IPSS or MOH), and the volume of outpatient visits.⁴⁰

³⁷ Used beds is a measure of output, not capacity.

³⁸ The average cost function estimated was

$$\ln ACPD = \ln a_0 + a_1 \cdot \ln SCALE + a_2 \cdot \ln OCR + a_3 \cdot \ln ALS + a_4 \cdot \ln TOPPD + a_5 \cdot \ln SAT + a_6 \cdot PHD + u$$

where the variables are as defined in the text and u is a random error term.

³⁹ This recommendation takes no account of location of demand, however.

⁴⁰ Dor estimated quadratic average cost functions of the following form:

$$AC_k = b_0 + b_1 \cdot F_k + b_2 \cdot (F_k)^2 + C' \cdot Z_k + v_k$$

where F denotes patient flow (after Feldstein, 1967), Z is a vector of case mix and technology variables, b_0 - b_2 are coefficients, C is a vector of coefficients, v is a random error term, and k is an index denoting the cost category (k = all cost categories; labor only; goods only; and services only).

Based on his average cost equation, Dor calculated an analytical expression for the optimal patient flow, i.e., the flow at which hospital average cost is minimum.⁴¹ Using OLS and weighted least squares (WLS), he estimated separate average cost functions for all cost categories combined as well as for each of his three cost categories separately (the categories were labor, goods, and services). He found that hospital average cost decreased with service intensity (although at a decreasing rate). This means that average cost of hospitalizations went down if the number of hospitalizations increased, or the number of beds decreased, or both. He computed an optimal flow value of 4.2 (admissions per bed per month) that was above the sample mean but below the actual flow of several hospitals in the sample. He thus concluded that the average cost curve was U-shaped. He also found that the aggregate number of hospital beds of both institutions should be reduced to move toward a minimum average cost point. Finally, he observed that, other things being equal, IPSS hospitals exhibited higher average costs than MOH institutions.

Bitran and Dunlop (1989) studied the determinants of hospital costs using a sample of 15 government hospitals in Ethiopia, with one to three annual observations for each, for a total of 38 observations. The study involved estimating a total cost function using OLS with a functional form similar to that introduced by Grannemann et al (1986) (see Section 4).⁴² Marginal costs of inpatient and outpatient care, average incremental cost, product-specific economies of scale, and economies of scope measures were analytically derived and evaluated from the estimated cost function.

The authors found that the hospitals in the sample were operating under constant returns to scale for patient days, laboratory exams, and deliveries. Economies of scope were found between first outpatient visits and inpatient days, signaling an economic advantage to the joint production of inpatient and outpatient care. The number of hospital beds affected total hospital costs in a statistically significant and economically important way. Input price proxy variables performed poorly in the regressions.

⁴¹ This is done by differentiating the AC equation with respect to F and equating it to zero, which yields $F^* = -b_1/(2 \cdot b_2)$.

⁴² The exact functional form used by Bitran and Dunlop was

$$\ln C = m_0 + m_1 \cdot BEDS + \sum_i (a_i \cdot \ln P_i) + f(Y) + \epsilon$$

where C is total annual cost, BEDS is the total number of beds in the facility, P_i is the price of the i^{th} production input, \ln denotes natural logarithm, $f(Y)$ is a function of various hospital products and outputs, and ϵ is a random error. In the absence of input prices, two alternative variables, that can be viewed as input price proxies, were used: P_1 = (number of physicians)/(total personnel) and P_2 = distance (in miles) between the hospital and the country's capital of Addis Ababa. P_1 captures possible differences in the average price of labor (those facilities with a higher percentage of physicians to total labor should have higher average wages); P_2 was intended to capture differences in input prices arising from differential transportation costs from the main distribution points in the capital city. The precise form of $f(Y)$ was

$$f(Y) = b_{11} \cdot IP + b_{12} \cdot OP + b_{13} \cdot DELIV + b_{14} \cdot XRAY + b_{15} \cdot SURG + b_{16} \cdot LAB + c_{11} \cdot IP^2 + c_{12} \cdot OP^2 + d_{11} \cdot IP \cdot OP$$

where IP is number of inpatient admissions, OP is number of outpatient visits, DELIV is number of deliveries, XRAY is number of X-rays performed, SURG is number of surgical interventions performed, and LAB is number of laboratory tests done.

A limitation of this type of analysis is that cost minimization cannot be assumed a priori. The estimated coefficients may be biased estimators of the minimum cost coefficients. As a consequence, the associated measures about technology (e.g., economies of scale and scope), derived from the estimated cost function, may also be biased (see discussion on this point under Eakin and Kniesner (1988) in Section 4.5). Another limitation of the work of these authors is that they used as an independent variable hospital expenditures, as reported by the Ethiopian government. This figure probably underestimated the actual resources used in production, because it did not include possible donor gifts and the depreciation of capital. An additional possible limitation is that reported expenditures may depart from the actual social cost of the resources used in hospital operations (see discussion of this point in Section 4). Finally, the sample of hospitals used by Bitran and Dunlop was bi-modal with respect to hospital size. Estimation of cost from the pooled sample of small and large hospitals could be a shortcoming if the true cost relations differed between the different size facilities. Unfortunately, the small number of observations precluded separate estimation of cost functions for small and large hospitals.

Wouters (1990) studied the costs and efficiency of a sample of 42 private and public health facilities in Ogun State, Nigeria. The sample included a heterogeneous range of facilities, including comprehensive health centers, primary health care clinics, maternities, health clinics, and dispensaries. Wouters analyzed costs and efficiency, estimating a production and a cost function, and deriving associated measures of efficiency.

Technical efficiency was assessed by estimating a production function and deriving measures of the marginal product of health workers. The (rather peculiar) production function expressed the number of outpatient visits (V) as a function of the number of inpatient admissions (IN), the number of patients receiving drugs (DRGNUM86), the proportion of high- and low-level health workers (HHW and LHW, respectively), and the presence and number of beds (BEDSDUMMY and BEDS, respectively).⁴³ To assess whether cost minimization was taking place, estimates of the relative marginal product of high- and low-level health workers were compared with the relative price of each category of labor, in both public and private facilities. She found that public facilities employed too many low-level health workers. In contrast, private providers were found to be using a near optimal (i.e., cost minimizing) mix of high- and low-level health workers.

The cost function estimated by Wouters included as independent variables the volume of outpatient visits and inpatient admissions, the proportion of patients obtaining drugs, the wages of

⁴³ The production function was a generalized translog Cobb-Douglas bi-product function, as follows

$$\ln V = a_0 + a_1 \cdot \frac{(IN^\Lambda - 1)}{\Lambda} + a_2 \cdot \ln DRGNUM86 + a_3 \cdot \ln HHW + a_4 \cdot \ln LHW \\ + a_5 \cdot BEDSDUMMY + a_6 \cdot \left[\frac{(BEDS^\Lambda - 1)}{\Lambda} \cdot BEDSDUMMY \right]$$

where \ln denotes natural logarithm and Λ is a parameter of the Box Cox metric. This generalized formulation allows for zero inpatient and bed levels.

both kinds of health workers, the availability and number of beds, and an efficiency index.⁴⁴ She found that the efficiency variable was insignificant and thus concluded that departures from cost minimization had little effect on expenditures. She also found that marginal costs were less than average costs and thus concluded that the facilities in her sample exhibited increasing returns to scale for both admissions and outpatient visits. With regard to economies of scope, she concluded that there were no apparent advantages to the joint production of inpatient and outpatient care.

Lewis et al. (1990) conducted a study of costs, efficiency, and quality of care at Aybara Hospital, a 271-bed government-run facility in the Dominican Republic. To compute costs, the authors tracked a selected sample of patients as they moved through the hospital, recording and pricing the services provided to them.⁴⁵ Where possible, non-labor prices were obtained directly from the supplier. Labor costs were computed by measuring actual staff time devoted to medical procedures and multiplying it by the worker's actual wage (converted to an hourly basis). The total cost of a procedure was computed as the sum of the variable (or direct) and allocated fixed costs. Variable costs were broken into labor, drugs, ancillary services, and consumables. Fixed costs included allocated overhead and the depreciation of buildings, equipment, and other fixed assets. Quality of care was measured in two ways: by assessing the appropriateness of the qualifications of the medical staff involved in care and by comparing actual diagnostic and treatment practices and services delivered with medical norms of care.

By extrapolating labor use from the sample to all services and patients in the hospital, Lewis et al. concluded that only 12% of the medical labor contracted by the hospital could actually be accounted for. Although no allowances were made for down time by the medical staff, this striking result signals a major inefficiency in the operations of Aybara Hospital. With regard to quality of care, major departures were found between the expected cost of meeting diagnostic and treatment

⁴⁴ The cost function was a generalized Cob-Douglas bi-product function:

$$\ln Rcost = b_0 + b_1 \ln V + b_2 \frac{(IN^{\Lambda} - 1)}{\Lambda} + b_3 \ln DRGPCT86 + b_4 \ln Index + b_5 \ln WagHHW + b_6 \ln WagLHW + b_7 \cdot BEDSDUMMY + b_8 \left[\frac{(BEDS^{\Lambda} - 1)}{\Lambda} \cdot BEDSDUMMY \right]$$

where Rcost is recurrent expenditures of the facility, the other variables are as defined in the preceding footnote, and Index is the efficiency index as proposed by Goldman and Grossman (1983):

$$Index = \left| \frac{\frac{MP_{LHW}}{WAGE_{LHW}}}{\frac{MP_{HHW}}{WAGE_{HHW}}} - 1 \right|$$

where MP denotes marginal product of health worker labor and $|\cdot|$ denotes absolute value.

⁴⁵ The sample consisted of all emergency patients during a one-week period; a sample of people consulting on an ambulatory basis during that week; and inpatients admitted to five (three surgical, two ophthalmology) of the hospital's 18 wards during a two-week reference period. The exercise also included 23 surgical interventions consisting of wounds, appendicitis, cataract, hysterectomy, and hernia.

norms and the cost of diagnostic and treatment services actually provided; the drugs dispensed and the tests done represented about 10 percent of the costs implied by the norms.

The study by Lewis et al. is one of the few that carefully examines resource use and technical quality of care in a developing-country hospital. The information the study provides is useful for hospital staff as well as government decision makers. A drawback of the study, however, is its elevated cost, arising from the need to price, and monitor carefully, through time and motion studies, resource use. To retain the method's advantages yet limit its cost, it has to be applied to a narrow set of interventions, and for a small, but statistically significant sample of patients, as was done by Lewis et al. and Pineault et al. (1985) (see Section 4).

Shepard et al. (1991) conducted a cost-effectiveness study of surgery in intermediate health units (IHUs) in Cali, Colombia. In an effort to reduce the high costs of hospital care, IHUs were created as intermediary settings between primary care facilities and hospitals. Two facilities were chosen for the comparative study: a 20-bed IHU, which performed about 1,100 operations of all types in 1988, and a 127-bed secondary hospital, which performed 3,500. Inguinal herniorrhaphy was selected as the intervention on which to base the study because of its high frequency, its moderate degree of technical complexity, and the existence of standardized indices of surgical risk in both facilities.

The measures of effectiveness included complication rates, patient satisfaction, and duration of post-operative disability. Complication rates were assessed through the eighth post-operative day by analyzing medical records by trained health professionals. Patient satisfaction was measured through two surveys which assessed the patient's period of convalescence. The first survey took place one week after the intervention, when the patient returned for his or her surgical follow-up visit (if the patient did not return, attempts were made to contact him or her at home). The second patient survey took place in his or her home four to seven months after the surgery.

Costs of care were measured from the patient's admission through the eighth post-operative day, thus excluding the costs of diagnostics tests prior to admission. Both direct and indirect costs of care were accounted for. Direct costs consisted of medical supplies, general supplies, personnel, and depreciation of medical equipment. The latter three direct costs were allocated to a single hernia repair based on the ratio of the intervention's time to the total time of use of the operating room in one year.

The study found that the rates of surgical and anesthesia complications were higher in the hospital (three complications) than in the IHU (none); patient satisfaction was higher in the IHU, as measured by the promptness with which patients were able to return to work after surgery. IHUs were found to use a mix of personnel that was less intensive in high cost professionals. The average cost of herniorrhaphy was \$39.12 at the IHU and \$148.76 at the hospital. Average out-of-pocket prices to patients were \$16.53 at the IHU versus \$38.96 at the hospital. When patient drug expenditures were added, average patient payments represented 30 and 48 percent of costs in the IHU and the hospital, respectively. The authors also found that economic efficiency could be improved in both settings with increased utilization, revealing under-utilization of fixed resources

at current demand levels. This finding has policy relevance only if there is unsatisfied demand, that is, if there are hernia patients who need surgery but do not get it.

5.2 SUMMARY AND CONCLUSIONS

Despite a large and growing body of literature on the measurement of health facility costs in developing countries, the literature on the measurement of efficiency is scant. While data limitations are undoubtedly the basis for the lack of research in this area, the limited volume of work may be explained largely by the fact that measuring efficiency is intrinsically much harder than measuring costs.⁴⁶

Of the seven studies reviewed, only one (Wouters, 1990) used data that were not from hospitals. In addition, all seven studies included government facilities in their sample, while only one of them (Wouters, 1990) included both private and public providers.

The emphasis on hospital efficiency research in developing countries coincides with that in the developed world, because hospitals account for the largest share of health care costs. The emphasis on government hospital efficiency, on the other hand, can be explained because either or both:

- (1) Governments tend to keep information on utilization and costs—however inaccurate—in a uniform way, whereas private providers generally do not; or
- (2) The search for health care financing and delivery reform has focused on gauging and improving productive efficiency of the public sector.

The lack of comparative studies of productive efficiency between government and private providers is surprising, in light of the growing—yet empirically unsupported—pressures on the part of experts and donors to promote public divestment of curative care production in favor of a growing private participation. The reasons that lead private providers to behave more efficiently than government institutions in other fields (e.g., agriculture) are likely to prevail in the health sector. Yet, the existing body of research appears to be too weak to support empirically the reforms so strongly promoted.

The following section proposes three studies of health provider efficiency in developing countries. These studies seek to:

- (1) Test existing techniques for the measurement of efficiency in the developing-country setting; and
- (2) Gain further empirical information about the levels and differences in efficiency between government and private providers.

⁴⁶ For a review of studies of hospital costs in developing countries see Barnum and Kutzin (1990).

The two objectives will be pursued in a balanced way in the hospital sector as well as in the case of ambulatory facilities.

6.0 PRELIMINARY RESEARCH DESIGN FOR PHASE TWO FIELD WORK

This paper has described techniques for the measurement of technical and economic efficiency, methodological problems associated with these measurements, and possible solutions. It has also provided a selective review of efficiency studies done in developed and developing countries.

This final section presents the research design for three studies on efficiency measurement. The first two studies are proposed in the context of a comparative study of hospital efficiency in Ecuador. The third study seeks to measure efficiency in a sample of public and private health care facilities in Senegal. The research design of these studies is described next.

6.1 A COMPARATIVE STUDY OF EFFICIENCY IN GOVERNMENT AND PRIVATE INPATIENT FACILITIES IN ECUADOR

USAID/Quito has recently asked HFS to submit a research proposal to assess the efficiency of five different types of inpatient providers in Ecuador. They are the Ministry of Public Health, the Junta de Beneficencia, the Social Security Administration, the private sector, and the military. The study sample would consist of five hospitals, one of each type.

This section proposes two existing methods to assess hospital efficiency:

- (1) The use of Data Envelopment Analysis (DEA) to measure overall hospital technical efficiency, as was done by Sherman (1981) for a group of six Massachusetts hospitals (see Section 4.6); and
- (2) The assessment of efficiency for a small sample of interventions, controlling for quality of care, as done by Pineault et al. (1985).

Separate research designs for these two methods follow. While there would be some economies if both studies were performed simultaneously, particularly at the data collection stage, the studies are proposed as two discrete activities. They can easily be combined as a single activity, however, if so desired by USAID/Quito.

6.1.1 Assessment of Overall Hospital Technical Efficiency Using the DEA Technique

6.1.1.1 Study Goals and Objectives

The goals of this study are to:

- (a) Measure overall technical efficiency in sample hospitals and
- (b) Make recommendations for improving health system efficiency.

The objectives of the study are to:

- (a) Compare technical efficiency among hospitals;
- (b) Make policy, technical, and managerial recommendations to improve efficiency in all five facilities; and
- (c) Demonstrate the use in the developing-country setting of existing methods for measuring overall technical efficiency.

6.1.1.2. Method and Workplan

The first study task will be to hold a technical and planning meeting in the U.S. with the participation of a consultant familiar with the use of the DEA technique to measure hospital efficiency. The DEA method requires information about hospital outputs and production inputs (see Section 4.6). Therefore, the next step will be to assemble a panel of hospital experts and managers from the five study facilities to define all relevant hospital outputs and inputs. At this meeting the notion of hospital technical efficiency will be discussed and the basic principles behind the DEA technique explained. The policy relevance of the study will be established. Data collection instruments will then be developed and tested. These instruments will be used to gather detailed information about output and input levels.

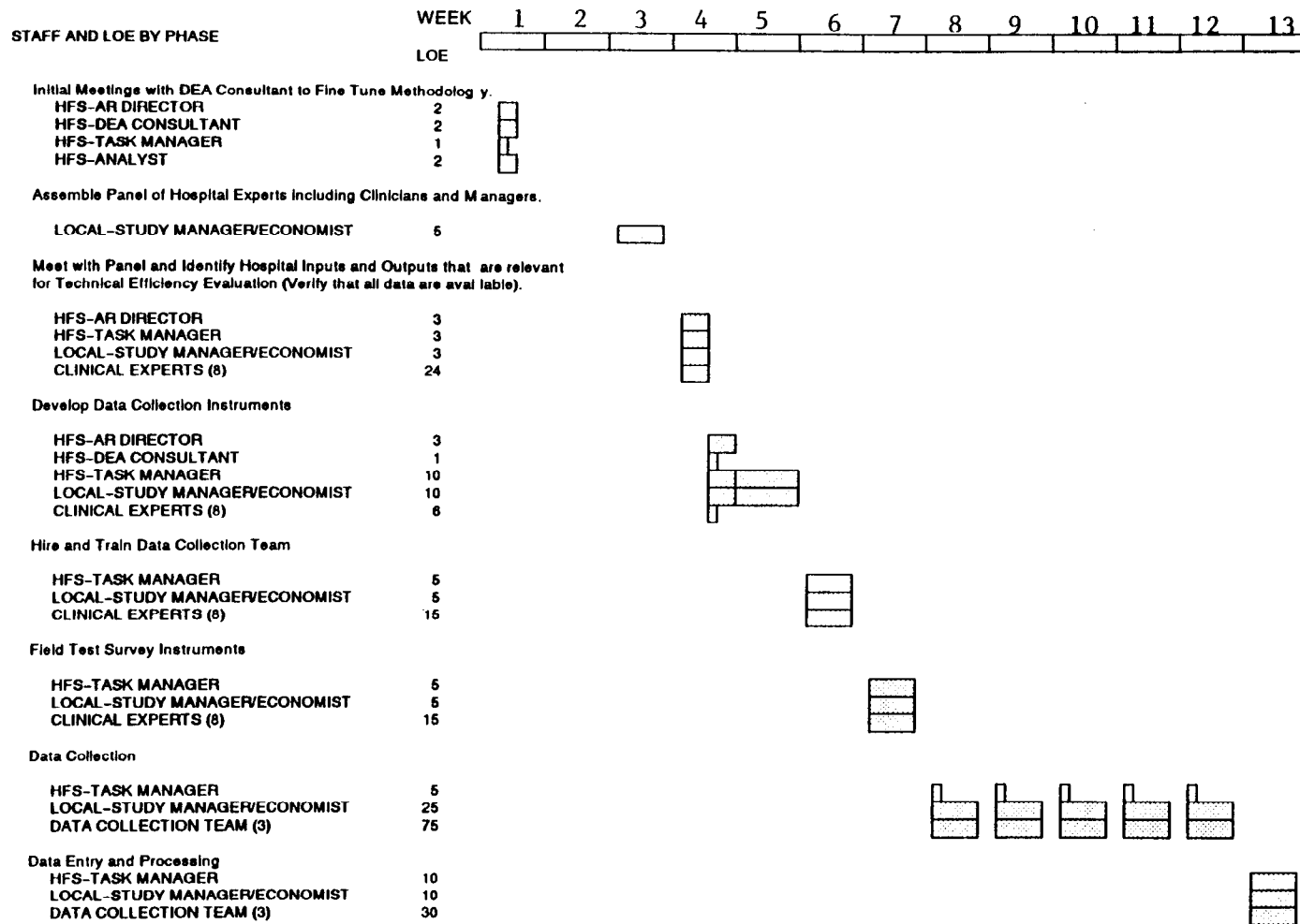
A data collection team will be assembled and trained both to gather data and to put them in the form required by the DEA linear programming algorithm. The DEA technique will then be used to compute a hospital-specific index of technical efficiency (ITE). The most technically efficient hospital(s) in the sample will be assigned an ITE of 1.0 while the others will obtain indices between 0 and 1. For the relatively inefficient hospitals, various simulations will be made through DEA to explore how alternative management measures would affect technical efficiency in those facilities. Measures that improve technical efficiency will be those that increase the facility's ITE.

A policy-oriented report will be written and circulated among managers and clinicians from the five study facilities. A two-day workshop will then be held in Quito to discuss the study results and recommendations. The outcomes of this workshop will be incorporated in the final report.

Exhibit 12 presents a tentative Gantt chart for the study. It is anticipated that the study would take place over a period of 29 weeks, or about 7 months.

Gantt Chart

LOE DISTRIBUTION

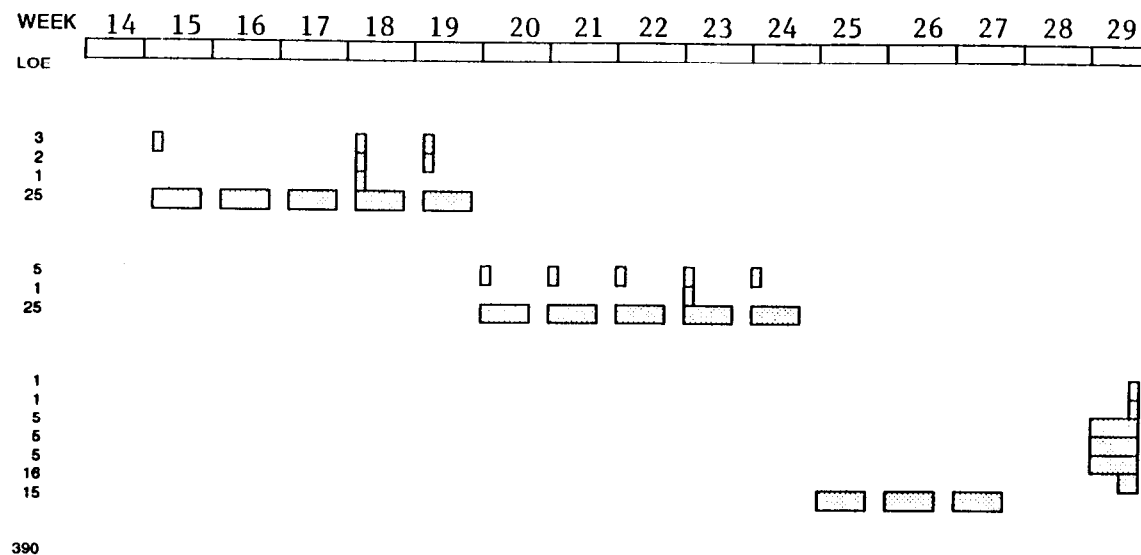


STAFF AND LOE BY PHASE

HFS-AR DIRECTOR
HFS-DEA CONSULTANT
HFS-TASK MANAGER
HFS-ANALYST

HFS-AR DIRECTOR
HFS-TASK MANAGER
HFS-ANALYST

HFS-AR DIRECTOR
HFS-TECH DIRECTOR
HFS-TASK MANAGER
HFS-ANALYST
LOCAL-STUDY MANAGER/ECONOMIST
CLINICAL EXPERTS (8)
TRANSLATOR



6.1.1.3 Level of Effort

The estimated level of effort for the study is summarized in Exhibit 13. The study is expected to require approximately 390 person-days or about 18 person-months. An estimate of the study labor costs is also provided in the exhibit. The study will be managed by an HFS staff member in conjunction with a local economist. A panel of about eight clinical experts will be recruited as technical advisors. The data collection staff will be either hospital managers or accountants, or economists familiar with hospital administration. The analysis of the data will be done by an HFS economist with input from both the HFS task manager and the local task manager.

6.1.2 Assessment of Economic Efficiency for a Selected Sample of Health Interventions

6.1.2.1 Study Goals and Objectives

The goals of this study are to:

- (a) Assess and compare economic efficiency for a group of five Ecuadorian hospitals for a sample of two frequently performed medical interventions and
- (b) Recommend measures for improving efficiency.

The objectives of the study are to:

- (a) Estimate the costs to the hospital of performing these interventions as well as the costs to the patients of undergoing care;
- (b) Assess clinical outcomes and patient satisfaction with the services;
- (c) Compare the above measures among hospitals;
- (d) Make policy, technical, and managerial recommendations to improve efficiency in all five facilities; and
- (e) Develop and use a methodology for measuring technical and economic efficiency in developing-country hospitals.

Study of Hospital Efficiency in Ecuador Using DEA Technique **Estimated Level of Effort**

ECUADOR (A)

	RATE	PERSON DAYS	AMOUNT	PHASE										
				1	2	3	4	5	6	7	8	9	10	11
HFS-SENIOR STAFF														
HFS-AR DIRECTOR	832	17	14,144	2		3	3					3	5	1
HFS-TECH DIRECTOR	832	1	832											1
HFS-DEA CONSULTANT	832	5	4,160	2			1					2		
SUBTOTAL		23	19,136											
HFS-MIDLEVEL STAFF														
HFS-TASK MANAGER	450	46	20,700	1		3	10	5	5	5	10	1	1	5
HFS-ANALYST	450	57	25,650	2								25	25	5
SUBTOTAL		103	46,350											
LOCAL-SENIOR STAFF														
LOCAL-STUDY MANAGER/ECON	150	68	10,200		5	3	10	5	5	25	10			5
CLINICAL EXPERTS (8)	150	76	11,400			24	6	15	15					16
TRANSLATOR	150	15	2,250											15
SUBTOTAL		159	23,850											
LOCAL-FIELD AND SUPPORT STAFF														
DATA COLLECTION TEAM (3)	30	105	3,150							75	30			
CLERK/SECRETARY	30	0	0											
SUBTOTAL		105	3,150											
TOTAL		390	92,486	7	5	33	30	25	25	105	50	31	31	48

6.1.2.2 Method and Workplan

The study will begin with meetings with hospital medical directors and managers to discuss the study goals and objectives. A committee of clinical experts will be appointed and two relatively uncomplicated medical interventions that are frequently performed in all facilities will be selected. Examples of such interventions are: normal deliveries, appendectomies, tonsillectomies, and hernia repairs. The clinical experts will agree on standards of treatment for the two interventions.

Methods for assessing the interventions' cost to the facilities and patients, patient satisfaction, and clinical outcomes will be developed. Facility treatment costs will be assessed by measuring and costing out the resources used directly in care as well as by allocating indirect costs to care. A sampling frame will be established to select a subset of patients undergoing each of the two interventions in all five facilities. Data collection for resource use will be done by tracking down all resources devoted to care, from the time the patient enters the facility to the time he or she is discharged. Forms and questionnaires will be developed and used for this purpose. Input prices as well as indirect costs will be gathered through facility-specific questionnaires.

Departures from the norms of care will be measured by comparing actual practices with standards of treatment. Assessment of clinical outcomes will be done by physician consultants through analysis of medical charts, examination of patients, and patient follow-up interviews after discharge. Patient satisfaction will be measured through interviews, which will seek information about patient perception of quality and attitude of staff, cleanliness of the premises, waiting time, post-operative discomfort, post-operative follow-up visits, and overall recovery.

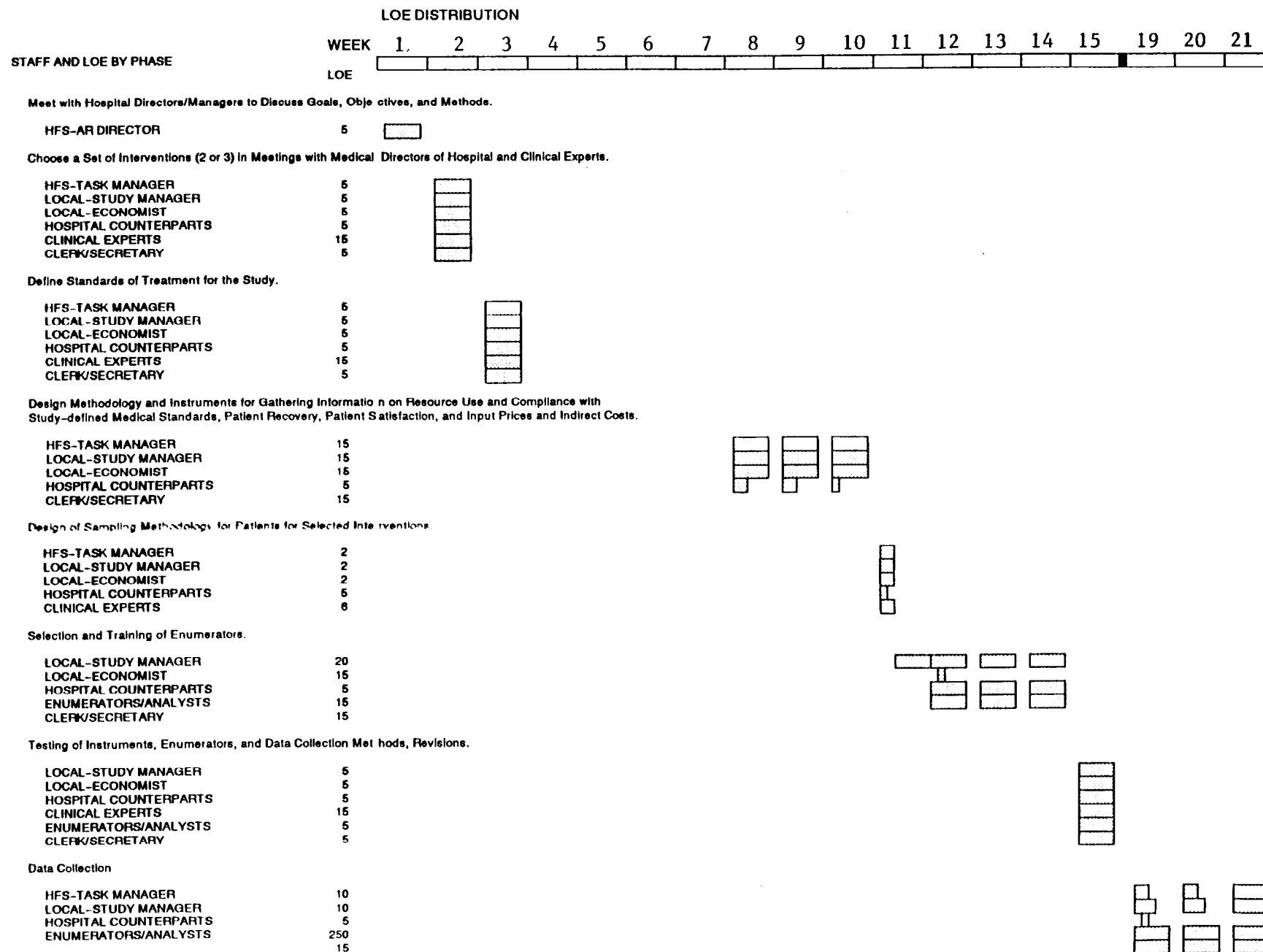
Economic efficiency will be assessed by combining overall treatment costs with information about patient quality perceptions, compliance with norms of care, and clinical outcome. The study findings will be included in a report to be shared with hospital managers and medical staff, as well as with USAID/Quito. A one-day workshop with the above participants will be conducted in Quito to discuss the study findings and recommendations. Advice will include possible management measures for improving efficiency. The outcome of the discussions at the workshop will be incorporated into the final report.

The study is expected to take place over a period of 38 weeks or approximately nine months, as is shown in the Gantt chart of Exhibit 14.

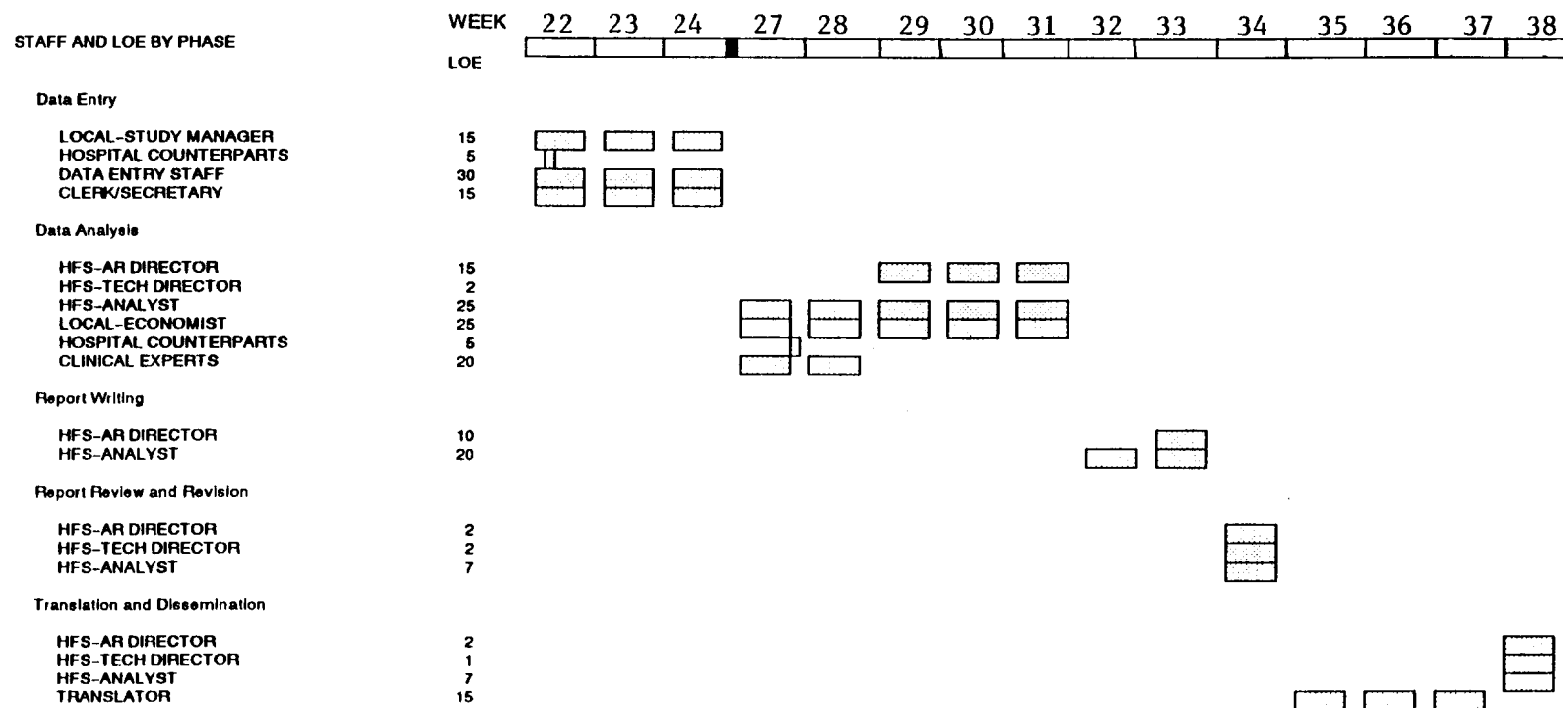
Study of Hospital Efficiency in Ecuador for Selected Sample of Interventions

Gantt Chart

ECUADOR (B) Assessment of Economic Efficiency for a Selected Sample of Health Interventions



ECUADOR (B)



6.1.2.3 Level of Effort

The study is expected to rely heavily on the use of local consultants, as is shown in Exhibit 15. These will include a study manager, an economist, possibly one or more accountants or managers, clinical experts, enumerators, data entry staff, and a study clerk. The following table summarizes the personnel requirements, including a rough estimate of the study labor costs. The study is expected to require about 790 person-days, or 36 person-months, with labor costs of approximately \$120,000. Personnel rates are estimates and are fully loaded.

Study of Hospital Efficiency in Ecuador for Selected Sample of Interventions **Estimated Level of Effort**

ECUADOR (B)

	RATE	PERSON DAYS	AMOUNT	PHASE												
				1	2	3	4	5	6	7	8	9	10	11	12	13
HFS-SENIOR STAFF																
HFS-AR DIRECTOR	832	34	28,288	5									15	10	2	2
HFS-TECH DIRECTOR	832	5	4,160												2	3
SUBTOTAL		39	32,448													
HFS-MIDLEVEL STAFF																
HFS-TASK MANAGER	450	37	18,650		5	5	15	2			10					
HFS-ANALYST	450	59	26,550										25	20	7	7
SUBTOTAL		96	43,200													
LOCAL-SENIOR STAFF																
LOCAL-STUDY MANAGER	75	77	5,775		5	5	15	2	20	5	10	15				
LOCAL-ECONOMIST	150	72	10,800		5	5	15	2	15	5			25			
HOSPITAL COUNTERPARTS	75	45	3,375		5	5	5	5	5	5	5	5	5			
CLINICAL EXPERTS	150	71	10,650		15	15		6		15			20			
TRANSLATOR	150	15	2,250													15
SUBTOTAL		280	32,850													
LOCAL-FIELD AND SUPPORT STAFF																
ENUMERATORS/ANALYSTS	30	270	8,100						15	5	250					
DATA ENTRY STAFF	30	30	900									30				
CLERK/SECRETARY	30	75	2,250		5	5	15		15	5	15	15				
SUBTOTAL		375	11,250													
TOTAL		790	119,748	5	40	40	65	17	70	40	290	65	90	30	11	27

6.2 A COMPARATIVE STUDY OF EFFICIENCY IN GOVERNMENT AND PRIVATE INPATIENT FACILITIES IN SENEGAL

USAID/Dakar has asked HFS to perform a study of health facility costs and efficiency using a sample of about 40 Ministry of Health facilities, including three hospitals, 18 health centers, nine maternities, 34 health posts, and 34 health huts. Although USAID is primarily interested in issues pertaining to MOH efficiency, HFS will seek to include in the study a sample of private sector providers. The field work will begin in early February 1993, with a visit to Senegal by an HFS staff member to interview local study counterparts and to plan the study logistics. USAID/Dakar will fund all the study costs associated with the analysis of MOH costs and efficiency and health committee performance. The incremental costs associated with the study of private facility costs and efficiency, as well as a comparison of costs and efficiency between MOH and private providers, will be borne by HFS.

6.2.1. Goals and Objectives

The study goals are to:

- (a) Improve knowledge about the levels and determinants of costs and efficiency for a sample of government and private health facilities; and
- (b) Recommend policy measures to improve the overall efficiency of the health system by making better use of government and private resources.⁴⁷

The study objectives are to:

- (a) Use information on government health facility costs to aid in budgeting and strategic financial planning;
- (b) Obtain information on health services utilization in a sample of government facilities; and
- (c) Recommend measures for improving the efficiency of government facilities.

6.2.2. Method and Workplan

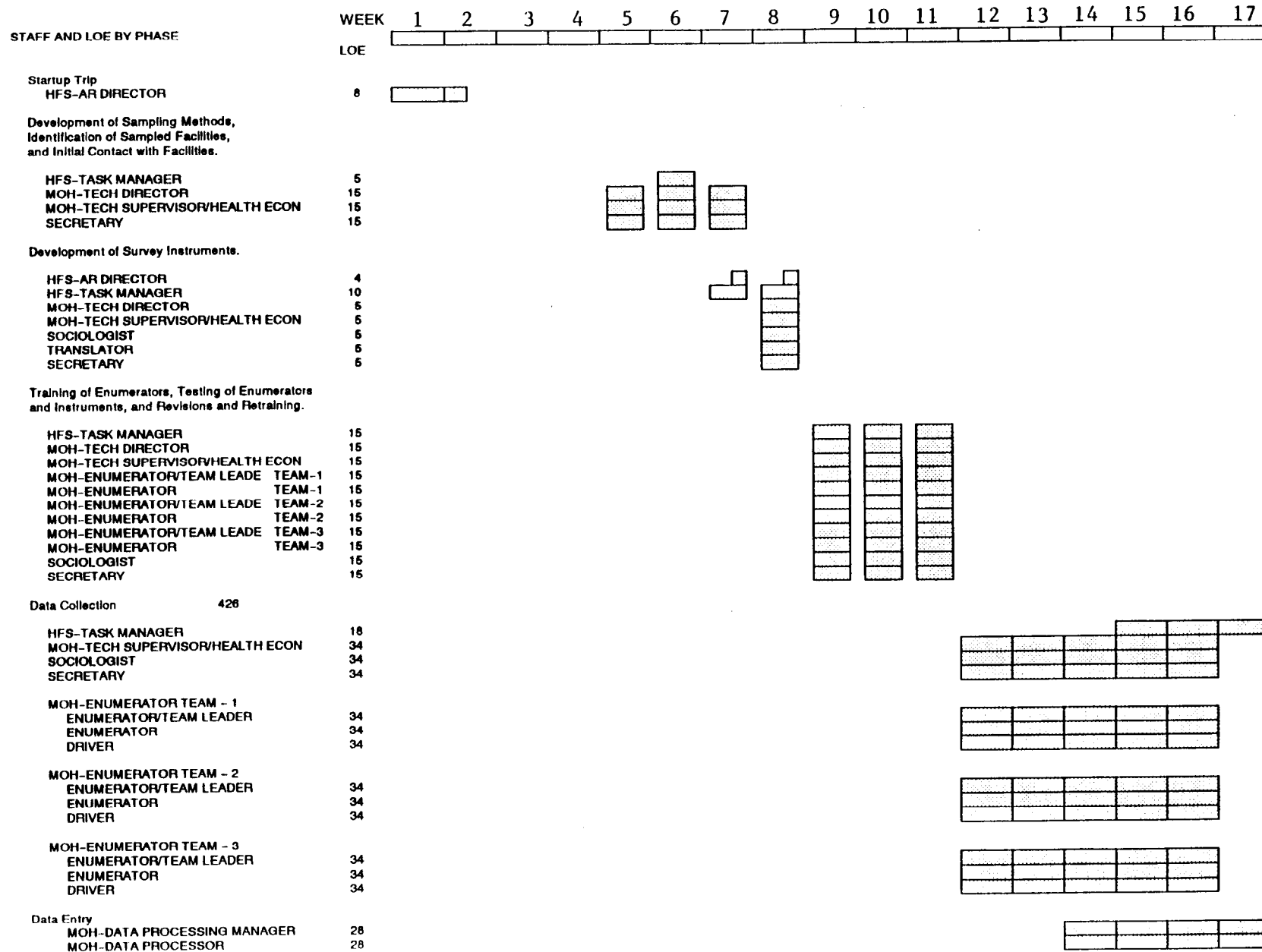
Following an initial visit to Dakar by HFS, the study will begin with the design of the sampling frame to select public and private ambulatory health facilities (see Exhibit 16). Hospitals, of which there are fewer than ten in the country, will be selected in a non-random fashion with the aim of obtaining geographic variation. Selected facilities will then be visited to seek their collaboration.

⁴⁷ The study of health committees has been requested by USAID/Dakar and is part of the purchase order.

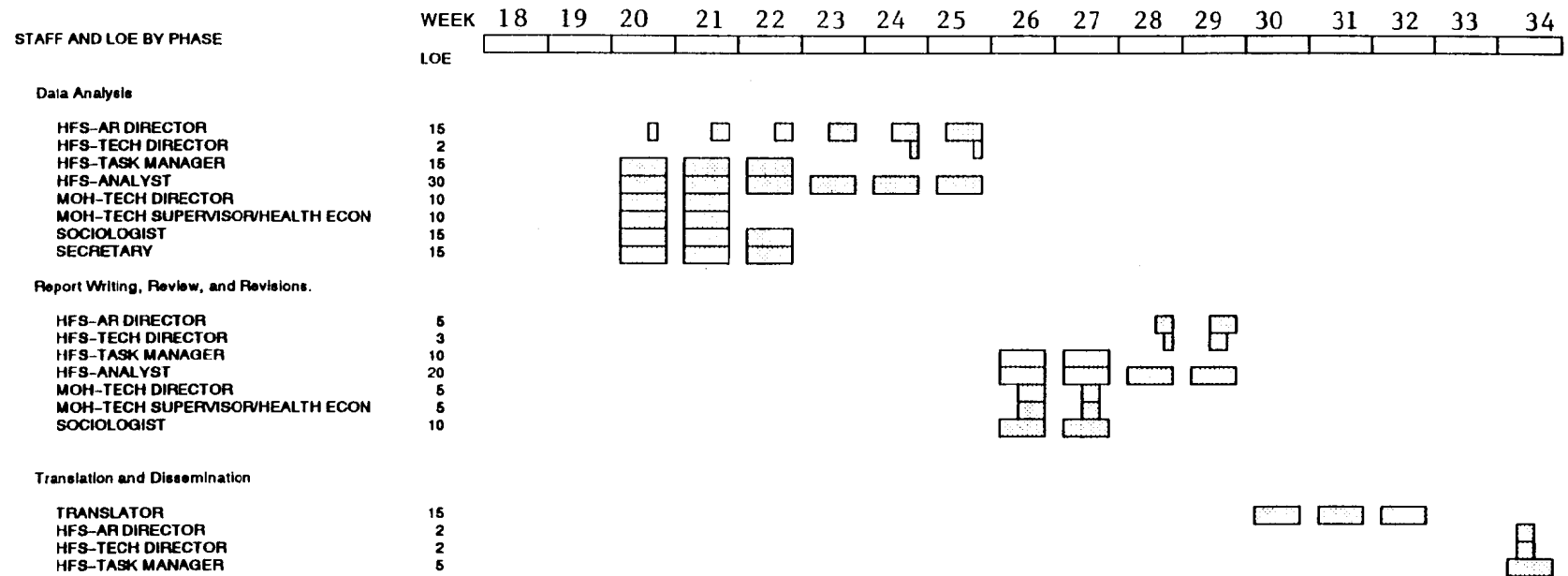
Study of Health Facility Costs and Efficiency and Health Committee Performance in Senegal

SENEGAL

LOE DISTRIBUTION



SENEGAL



Next, survey instruments will be developed and translated into French. The instruments will consist of facility questionnaires which will gather data on output levels, input prices, input quantities, facility costs, staff, and assets. The appropriate period from which facility data will be collected will be established. Enumerators will be hired and trained in data collection. The questionnaires, as well as the skills of the enumerators, will be field-tested. Necessary changes in the survey instruments will be made and additional personnel training will be provided, as needed. The study will then enter the data collection stage, to be performed through three teams of enumerators. Data entry will start two weeks after the beginning of the data collection phase. The facility data will be converted into ASCII files suitable for analysis through standard statistical packages such as SAS, SPSS, LIMDEP, and Lotus 1-2-3.

The analysis of health facility costs and efficiency will be done in the U.S. A report summarizing the methods, findings, and recommendations will be written in the U.S. and presented in Dakar.

The study of health committees will be conducted simultaneously with that of health facility costs and efficiency, as is shown in the Gantt chart of Exhibit 16. A sampling frame to select health committees will be developed. It is anticipated that a subset of the committees of the facilities surveyed will constitute the sample for the qualitative study of health committee performance.

Hospital efficiency will be measured using the DEA method in a fashion similar to that described for the study of hospital efficiency in Ecuador (see above).⁴⁸ Using the same technique in both studies will permit a comparison of results and an assessment of the advantages and disadvantages of this technique in different settings. With regard to ambulatory facilities, econometric estimation of cost and production functions will be performed. Flexible functional forms for cost and production functions, like those used by Eakin and Kniesner (1988) and Grannemann et al. (1986), will be adopted. The assumption of cost minimization will be tested.

The study will use cost and utilization information to illustrate how facility data can be used in budgeting. A computer model will be developed to illustrate how various fee levels would affect utilization, revenue, costs, and cost recovery performance, using alternative assumptions about demand response to price. The study is expected to take place over a six-month period.

⁴⁸The purchase order from USAID/Dakar does not include hospitals in the study sample. HFS will recommend that hospitals, both private and public, be added. The following Gantt chart and level of effort estimates do not include the activities and resources required to assess hospital efficiency. If hospitals do become part of the study, the research design will be modified accordingly.

6.2.3. Level of Effort

The estimated level of effort for this study is 938 person-days, or 43 person-months, as shown in Exhibit 17. Approximately one-half of this effort would be undertaken by HFS staff or HFS U.S. consultants. The study envisions heavy use of in-country personnel for the data collection and entry phase.

Study of Health Facility Costs and Efficiency and Health Committee Performance in Senegal
Estimated Level of Effort

SENEGAL

	<u>RATE</u>	<u>PERSON DAYS</u>	<u>AMOUNT</u>
HFS-SENIOR STAFF			
HFS-AR DIRECTOR	832	34	28,288
HFS-TECH DIRECTOR	832	7	5,824
SUBTOTAL		41	34,112
HFS-MIDLEVEL STAFF			
HFS-TASK MANAGER	450	78	35,100
HFS-ANALYST	450	50	22,500
SUBTOTAL		128	57,600
LOCAL-SENIOR STAFF			
MOH-TECH DIRECTOR	75	50	3,750
MOH-TECH SUPERVISOR/HEALTH ECO	150	84	12,600
SOCIOLOGIST	75	79	5,925
TRANSLATOR	150	20	3,000
SUBTOTAL		233	25,275
LOCAL-FIELD AND SUPPORT STAFF			
MOH-ENUMERATORS/TEAM LEADER	30	249	7,470
MOH-DRIVER	30	147	4,410
MOH-DATA ENTRY STAFF	30	56	1,680
SECRETARY	30	84	2,520
SUBTOTAL		536	16,080
TOTAL		938	133,067

BIBLIOGRAPHY

- ANDERSON, David L. (1980) "A Statistical Cost Function Study of Public General Hospitals in Kenya." 14(2): 223-235, 1980.
- BARNUM, H., BARLOW, R., FAJARDO, L., PRADILA, A. (1980) "Resource Allocation Model for Child Survival", Oelgeschlager, Gunn, Hain: Cambridge, MA.
- BARNUM, H. (1990) China Health Sector Study, The World Bank, unpublished.
- BARNUM, Howard and Joseph KUTZIN. (1990) "Public Hospitals in Developing Countries: Resource Use, Costs and Financing." Population and Human Resources Department, The World Bank: Washington, D.C.
- BERMAN, Peter. (1989) "Cost Efficiency in Primary Health Care: Studies of Health Facilities in Indonesia." *Health Policy and Planning*, Volume 4, No. 4, pp. 316-322.
- BERMAN, P., M. Nadjib BROTOWASISTO, S. SAKAI, and A. GANI. (1989) "The Costs of Public Primary Health Care Services in Rural Indonesia." *Bulletin of the World Health Organization*, Volume 67, No. 6, pp. 685-694.
- BERMAN, Peter and Suomi SAKAI. (1987) "The Productivity of Manpower and Supplies in Public Health Services in Java." The Johns Hopkins University, Department of International Health: Baltimore, MD.
- BERWICK, Donald M., M.D. (1989) "Health Services Research and Quality of Care — Assignments for the 1990s." *Medical Care*, Volume 27, No. 8, pp. 763-771.
- BITRAN, Ricardo and David W. DUNLOP. (1989) "The Determinants of Hospital Costs — An Analysis of Ethiopia." WPS 249, *Policy, Planning and Research Working Papers*, Population, Health and Nutrition Department, The World Bank: Washington, D.C.
- BITRAN, Ricardo and Steven BLOCK. (1992) "Provider Incentives and Productive Efficiency." Health Financing and Sustainability Project, Abt Associates Inc.: Bethesda, MD.
- BITRAN, Ricardo, and Keith McINNES. (1993) "The Demand for Health Care in Latin America: Lessons Drawn from the Dominican Republic and El Salvador." Economic Development Institute of the World Bank: Washington, D.C. Forthcoming.
- BLADES, C.A., A.J. CULYER and A. WALKER. (1987) "Health Service Efficiency: Appraising the Appraisers — A Critical Review of Economic Appraisal in Practice." *Social Science and Medicine*, Volume 25, No. 5, pp. 461-472.

- BLIERSBACH, Christopher M. (1989) "Integrating Quality Assurance and Cost Accounting to Facilitate Management Efficiency." *QRB*, October, pp. 302-305.
- BROOK, Robert H., M.D., ScD. and Kathleen N. LOHR, Ph.D. (1985) "Efficacy, Effectiveness, Variations, and Quality — Boundary-crossing Research." *Medical Care*, Volume 23, No. 5, pp. 710-722.
- BROWN, Randall S., Douglas W. CAVES, and Laurits R. CHRISTENSEN. (1979) "Modelling the Structure of Cost and Production for Multiproduct Firms," *Southern Economic Journal*, Vol. 46, July, pp.256-273.
- BYRNES, Patricia and Vivian VALDMANIS. (1989) "Variable Cost Frontiers and Efficiency: An Investigation of Labor Costs in Hospitals. *Lecture Notes in Economics and Mathematical Systems Series*. Berlin: Springer-Verlag, 332, pp.89-109.
- CAVES, Douglas W. and Laurits R. CHRISTENSEN. (1980) "The Relative Efficiency of Public and Private Firms in a Competitive Environment: The Case of Canadian Railroads." *Journal of Political Economy*, Volume 88, No. 5, pp. 958-976.
- CAVES, Douglas W., Laurits R. CHRISTENSEN, and Michael W. TRETHEWAY. (1980), "Flexible Cost Functions for Multiproduct Firms." *The Review of Economics and Statistics*, August, pp.477-481.
- CHARNES, A., W.W. COOPER, E. RHODES. (1978), "Measuring the Efficiency of Decision Making Units", *European Journal of Occupational Research*, Vol. 2, pp. 429-444.
- COWING, T. G., A. G. HOLTMANN and S. POWERS. (1983) "Hospital Cost Analysis: A Survey and Evaluation of Recent Studies." *Advances in Health Economics and Health Services Research*, Vol. 4, pp. 257-303.
- COWING, Thomas G. and Alphonse G. HOLTMANN (1983) "The Multiproduct Short-Run Hospital Cost Function: Empirical Evidence and Policy Implications from Cross-Section Data." *Southern Economic Journal*, January, pp. 637-653.
- DOR, Avi, Ph.D. and Dean E. FARLEY, Ph.D. (1992) "The Cost and Quality of Hospital Care by Source of Payment: Inferences from a Multiproduct Cost Function." Presented at National Bureau of Economic Research, Summer Institute.
- DOR, Avi, Ph.D., Philip J. HELD, Ph.D., and Mark V. PAULY, Ph.D. (1992) "The Medicare Cost of Renal Dialysis: Evidence from a Statistical Cost Function." *Medical Care*, Vol. 30, No. 10, pp. 879-891.
- DOR, Avi. (1987) "Estimates of Cost Function for the Peruvian Hospital Sector: Some Policy Implications." The Urban Institute: Washington, D.C.

- EAKIN, B. Kelly. (1991) "Allocative Inefficiency in the Production of Hospital Services." *Southern Economic Journal*, Vol. 58, No. 1, July, pp. 240-248.
- EAKIN, B. Kelly and Thomas J. KNIESNER. (1988) "Estimating a Non-minimum Cost Function for Hospitals." *Southern Economic Journal*, January, pp. 583-597.
- EVANS, Robert G., Ph.D., and Geoffrey C. ROBINSON, M.D. (1980) "Surgical Day Care: Measurements of the Economic Payoff." *Canadian Medical Association Journal*, Volume 123, pp. 873-880.
- FARRELL, M. (1957) "The Measurement of Productive Efficiency." *Journal of the Royal Statistical Association*, Series A.
- FARRELL, M. and M. FIELDHOUSE. (1962) "Estimating Efficient Production Functions Under Increasing Returns to Scale." *Journal of the Royal Statistical Association*.
- FELDSTEIN, M.S. (1974) "Econometric Studies of Health Economics." In M. Intrilligator and D. Kendrick (eds.), *Frontiers of Quantitative Economics*, Amsterdam: North Holland.
- FRANCISCO, E.W. (1970) "Analysis of Cost Variations Among Short Term General Hospitals." In H. Klarman, ed., *Empirical Studies in Health Economics*, Johns Hopkins University Press, Baltimore, MD.
- FRANK, Richard G. and Carl A. TAUBE. (1987), "Technical and Allocative Efficiency in Production of Outpatient Mental Health Clinic Services." *Social Science and Medicine*, Volume 24, No. 10, pp. 843-850.
- FREDERIKSEN, Kirsten. (1989) "Towards Improved Management of Costs and Revenues as a Means Towards Self-Sustainability of Health Services in Haiti." Institute for Child Health, Haiti.
- FRIEDMAN, Bernard and Mark V. PAULY. (1983) "A New Approach to Hospital Cost Functions and Some Issues in Revenue Regulation." *Health Care Financing Review*, Volume 4, No. 3, pp. 105-114.
- GARNER, Paul, Jane THOMASON and Dayl DONALDSON. (1990) "Quality Assessment of Health Facilities in Rural Papua New Guinea." *Health Policy and Planning*, Volume 5, No. 1, pp. 49-59.
- GOLD, Marsha, M.C.P, M.P.H. (1979) "Hospital-Based Versus Free-Standing Primary Care Costs." *Journal of Ambulatory Care Management*, February, pp. 1-20.

- GOLDMAN, Fred and Michael GROSSMAN. (1983) "The Production and Cost of Ambulatory Medical Care in Community Health Centers." *Advances in Health Economics and Health Services Research*, Volume 4, pp. 1-56.
- GOLDSCHMIDT, Taaqov and Amiram GAFNI. (1990) "A Managerial Approach to Allocating Indirect Fixed Costs in Health Care Organizations." *Health Care Management Review*, Volume 15, No. 2, pp. 43-51.
- GOLLADAY, Frederick L., Marilyn E. MANSER, and Kenneth R. SMITH (1974) "Scale Economies in the Delivery of Medical Care: A Mixed Integer Programming Analysis of Efficient Manpower Utilization." *The Journal of Human Resources*, Volume IX, No. 1, pp. 50-62.
- GRANNEMANN, Thomas W., Randall S. BROWN, and Mark V. PAULY. (1986) "Estimating Hospital Costs — A Multiple-Output Analysis." *Journal of Health Economics*, Volume 5, pp. 107-127.
- GURURAJ, Vymutt J., M.D., Teresita A. LAUDE, M.D., Solomon V. RAJKUMAR, M.D., Raymond M. RUSSO, M.D. (1982) "Observations of the Efficiency of Pediatric Ambulatory Care in a Teaching Clinic." *The Journal of Ambulatory Care Management*, February, pp. 41-46.
- HFS PROJECT. HFS Applied Research Agenda. Health Financing and Sustainability Project. Abt Associates Inc.: Bethesda, MD.
- HORNBROOK, Mark C., Ph.D. (1982) "Review Article — Hospital Case Mix: Its Definition, Measurement and Use: Part I. The Conceptual Framework." *Medical Care Review*, Vol. 39, pp. 1-43.
- HORNBROOK, Mark C., Ph.D. (1982) "Review Article — Hospital Case Mix: Its Definition, Measurement and Use: Part II. Review of Alternative Measures," *Medical Care Review*, Vol. 39, pp. 73-123.
- JACOBS, Philip. (1974) "A Survey of Economic Models of Hospitals." *Inquiry*, Volume XI, pp. 83-97.
- LASSO, Hipólito Pabón. (1986) "Evaluating Hospital Performance Through Simultaneous Application of Several Indicators." *Pan American Health Organization Bulletin*, Volume 20, No. 4, pp. 341-357.
- LAVE, Judith R. and Lester B. LAVE. (1984) "Hospital Cost Functions." *Annual Review of Public Health*, Volume 5, pp. 193-213.

- LEE, A. James, Ph.D., Harry B. WOLFE, Ph.D., and Betty SWENSON, R.N., B.S. (1984) "Hospital-Based Versus Free-Standing Ambulatory Care Costs: A Comparison." *Journal of Ambulatory Care Management*, February, pp. 25-39.
- LEWIS, Maureen A., Margaret B. SULVETTA and Gerard LAFORGIA. (1990) "Economic Costs of Hospital Services in the Dominican Republic." The Urban Institute: Washington, D.C.
- LEWIS, Maureen A., Margaret B. SULVETTA and Gerard M. LAFORGIA. (1990) "Measuring Costs, Efficiency, and Quality in Public Hospitals: A Dominican Case." Internal Discussion Paper, Report No. IDP-090, The World Bank: Washington, D.C.
- McGUIRE, Alistair. (1987) "The Measurement of Hospital Efficiency." *Social Science and Medicine*, Volume 24, No. 9, pp. 719-724.
- NUTTING, Paul A., M.D., Barton R. BURKHALTER, Ph.D., Duane L. DIETRICK, Ph.D., and Edward F. HELMICK, MPH (1982) "Relationship of Size and Payment Mechanism to System Performance in 11 Medical Care Systems." *Medical Care*, Volume XX, No. 7, pp. 676-690.
- NUTTING, Paul A., M.D., Gregory I. SHORR, M.D., and Barton R. BURKHALTER, Ph.D. (1981) "Assessing the Performance of Medical Care Systems: A Method and Its Application." *Medical Care*, Volume XIX, No. 3, pp. 281-296.
- O'DONOVAN, Thomas R., Ph.D. (1979) "Cost of Services in Ambulatory Surgical Facilities." *The Journal of Ambulatory Care Management*, November, pp. 23-28.
- OVER, Mead, Jr. (1984) "Cost-Effective Integration of Immunization and Basic Health Services in Developing Countries: The Problem of Joint Costs." Boston University: Boston, Massachusetts.
- OVER, Mead, Jr. (1985) "The Effect of Scale on Cost Projections for a Primary Health Care Program in a Developing Country." *Social Science and Medicine*, Volume 22, No. 3, pp. 351-360.
- PAULY, Mark V. (1970) "Efficiency, Incentives and Reimbursement for Health Care." *Inquiry*, Volume 7, No. 1, pp. 114-131.
- PINEAULT, Raynald, M.D., Ph.D., André-Pierre CONTANDRIOPOULOS, Ph.D., Marie VALOIS, M.Sc., Marie-Lynn BASTIAN, B.A., and Jean-Marie LANCE, M.Sc. (1985) "Randomized Clinical Trial of One-Day Surgery — Patient Satisfaction, Clinical Outcomes, and Costs." *Medical Care*, Volume 23, No. 2, pp. 171-182.

- PRESCOTT, R.J., Constance CUTHBERTSON, Norma FENWICK, W.M. GARRAWAY, and C. V. RUCKLEY. (1978) "Economic Aspects of Day Care After Operations for Hernia or Varicose Veins." *Journal of Epidemiology and Community Health*, Volume 32, pp. 222-225.
- PUGLISI, Robert J., Cindy Lou PARKS, William J. BICKNELL. (1991) "Functional Expenditure Analysis for the Lesotho Ministry of Health." Draft, Boston University School of Public Health: Boston, MA.
- REGISTER, Charles A. and Edward R. BRUNING. (1986) "Profit Incentives and Technical Efficiency in the Production of Hospital Care." *Southern Economic Journal*, Volume 53, pp. 899-914.
- REINHARDT, U. (1972) "A Production Function for Physician Services." *The Review of Economics and Statistics*, Volume 60, pp. 55-66.
- RODRIGUEZ, Jorge and Jorge JIMENEZ. (1985) "Comparative Analysis of Productivity in Public and Private Chilean Hospitals." *Encuentro Anual de Economistas 1985*, Departamento de Economia, Universidad de Santiago de Chile.
- ROSENSTEIN, Alan H. (1991) "Health Economics and Resource Management: A Model for Hospital Efficiency." *Hospital and Health Services Administration*, Volume 36, No. 3, pp. 313-330.
- ROSKO, Michael D., Ph.D. (1990) "Measuring Technical Efficiency in Health Care Organizations." *Journal of Medical Systems*, Volume 14, No. 5, pp. 307-322.
- SCHMIDT, Peter and C. A. Knox LOVELL. (1980), "Estimating Stochastic Production and Cost Frontiers when Technical and Allocative Inefficiency are Correlated." *Journal of Econometrics*, Volume 13, pp. 83-100.
- SEAR, Alan M. (1991) "Comparison of Efficiency and Profitability of Investor-Owned Multihospital Systems with Not-for-Profit Hospitals." *Health Care Management Review*, Volume 16, No. 2, pp. 31-37.
- SHEPARD, Donald S., Julia WALSH, Wolfgang MUNAR, Laura ROSE, Rodrigo GUERRERO, Luis F. CRUZ, Guillermo REYES, Gail P. ORSOLANI, Carlos SOLARTE. (1991) "Cost-Effectiveness of Ambulatory Surgery in Cali, Colombia." Department of Population Sciences, Harvard School of Public Health: Boston, MA.
- SHERMAN, H.D. (1984) "Hospital Efficiency Measurement and Evaluation. Empirical Test of a New Technique." *Medical Care*, Vol. 22, No. 10.

- SHERMAN, H.D. (1981) "Measurement of Hospital Technical Efficiency: A Comparative Evaluation of Data Envelopment Analysis and Other Efficiency Measurement Techniques for Measuring and Locating Inefficiency in Health Care Organizations." Unpublished D.B.A. Thesis, Harvard University.
- SOULIER, Mary Ziebell and Don T. DECOSTER. (1982) "Productivity Versus Cost Control: Considerations for Health Care Managers." *Health Care Management Review*, Winter, pp. 15-20.
- TATCHELL, Michael. (1983) "Measuring Hospital Output: A Review of the Service Mix and Case Mix Approaches." *Social Science and Medicine*, Volume 17, No. 13, pp. 871-883.
- THORNGREN, Teresa A. and Thomas C. TINSTMAN. (1990) "Challenging New Roles for Medical and Professional Staff: Managing Costs and Quality of Care." *Topics in Health Care Financing*, Spring, pp. 1-11.
- TIMMER, C. Peter (1970) "On Measuring Technical Efficiency." *Food Research Institute Studies in Agricultural Economics, Trade, and Development*, Volume IX, No. 2, pp. 99-171.
- TODA, Yasushi (1976) "Estimation of a Cost Function When the Cost is not Minimum: The Case of Soviet Manufacturing Industries, 1958-1971." *The Review of Economics and Statistics*, Volume LVIII, No. 3, pp. 259-268.
- VALDMANIS, Vivian G., Ph.D. (1990) "Ownership and Technical Efficiency of Hospitals." *Medical Care*, Volume 28, No. 6, pp. 552-561.
- VARIAN, H. (1987) *Microeconomic Analysis*, Second Edition, W.W. Norton & Company.
- VITALIANO, Donald F. (1987) "On the Estimation of Hospital Cost Functions." *Journal of Health Economics*, Volume 6, pp. 305-318.
- WAN, Thomas T. H., Ph.D., Paul R. VANOSTENBERG, DDS, John J. SALLEY, DDS, Ph.D., David W. SINGLEY, MHA, John L. WEST, MHA. (1987) "A Cost and Production Analysis of Hospital Dental Care Programs." *Public Health Reports*, Volume 102, No. 5, pp. 512-522.
- WILSON, George W. and Joseph M. JADLOW (1982) "Competition, Profit Incentives, and Technical Efficiency in the Provision of Nuclear Medicine Services." *The Bell Journal of Economics*, Volume 13, No. 2 (Autumn), pp. 472-482.
- WOUTERS, Annemarie. (1990) "The Cost and Efficiency of Public and Private Health Care Facilities in Ogun State, Nigeria." The Johns Hopkins University School of Hygiene and Public Health, Working Paper No. 4: Baltimore, MD.

- WYSZEWIANSKI, Leon, J. William THOMAS, Bruce A. FRIEDMAN. (1987) "Case-Based Payment and the Control of Quality and Efficiency in Hospitals." *Inquiry*, Volume 24, pp. 17-25.
- YANKAUER, Alfred, M.D., MPH. (1983) "Lessons in Surgery for the Third World." *American Journal of Public Health*, Volume 73, No. 12, pp. 1359-1360.
- ZASTOWNY, Thomas R., Ph.D., Klaus J. ROGHMANN, Ph.D., and Acco HENGST, Ph.D. (1983) "Satisfaction with Medical Care: Replications and Theoretic Reevaluation," *Medical Care*, Volume XXI, No. 3, pp. 294-322.